Monitoring Energy Performance in Local Authority Buildings

Graeme Stuart

A thesis submitted in partial fulfilment of the requirements of De Montfort University for the degree of Doctor of Philosophy

March 2011

Institute of Energy and Sustainable Development

De Montfort University, Leicester

Abstract

Energy management has been an important function of organisations since the oil crisis of the mid 1970's led to hugely increased costs of energy. Although the financial costs of energy are still important, the growing recognition of the environmental costs of fossil-fuel energy is becoming more important. Legislation is also a key driver. The UK has set an ambitious greenhouse gas (GHG) reduction target of 80% of 1990 levels by 2050 in response to a strong international commitment to reduce GHG emissions globally.

This work is concerned with the management of energy consumption in buildings through the analysis of energy consumption data. Buildings are a key source of emissions with a wide range of energy-consuming equipment, such as photocopiers or refrigerators, boilers, air-conditioning plant and lighting, delivering services to the building occupants.

Energy wastage can be identified through an understanding of consumption patterns and in particular, of changes in these patterns over time. Changes in consumption patterns may have any number of causes; a fault in heating controls; a boiler or lighting replacement scheme; or a change in working practice entirely unrelated to energy management.

Standard data analysis techniques such as degree-day modelling and CUSUM provide a means to measure and monitor consumption patterns. These techniques were designed for use with monthly billing data. Modern energy metering systems automatically generate data at half-hourly or better resolution. Standard techniques are not designed to capture the detailed information contained in this comparatively high-resolution data. The introduction of automated metering also introduces the need for automated analysis.

This work assumes that consumption patterns are generally consistent in the short-term but will inevitably change. A novel statistical method is developed which builds automated event detection into a novel consumption modelling algorithm. Understanding these changes to consumption patterns is critical to energy management.

Leicester City Council has provided half-hourly data from over 300 buildings covering up to seven years of consumption (a total of nearly 50 million meter readings).

Automatic event detection pinpoints and quantifies over 5,000 statistically significant events in the Leicester dataset. It is shown that the total impact of these events is a decrease in overall consumption.

Viewing consumption patterns in this way allows for a new, event-oriented approach to energy management where large datasets are automatically and rapidly analysed to produce summary meta-data describing their salient features. These event-oriented meta-data can be used to navigate the raw data event by event and are highly complementary to strategic energy management.

Acknowledgements

This research was funded as an INREB/EPSRC industrial CASE studentship. The industrial partners were Leicester City Council and Pilkington Energy Efficiency Trust. I would like to thank Leicester Energy Agency for their support, especially Don Lack and Prakash Patel. Without whom this work would not have been possible.

Thanks also go to my supervisors Paul Fleming, Simon Taylor and Vic Hanby. It may be said that I am not the easiest student to supervise. During the long process of writing up my thesis my requests for guidance were received with patience and generosity. I also acknowledge the contribution of Peter Harris who first introduced me to CUSUM analysis and inspired me to develop the method further.

I also happily acknowledge the contributions of many IESD Colleagues who provided me with moral support and advice on how to keep sane during difficult times. An incomplete list includes Raul Ajmat, Iraklis Argyriou, Keith Baker, Katy Beadle, Neil Brown, Michael Coleman, Paul Cropper, Tony Dowsett, Denis Fan, Stephen Firth, Vasco Ferreira, Miaomiao He, Kate Irvine, Kerry Kirwan, Ivan Korolija, Kuskana Kubaha, John Mardaljavic, Greig Mill, Rob Liddiard, Selveraj Naicker, Divine Novieto, Giuseppe Pellegrini-Masini, Stephen Porritt, Andrew Reeves, Mark Rylatt, Ashish Shukla, Rob Wall, Andrew Wallace, Tom White and William Wright. Thanks to you all, sorry if I have missed anyone.

I would never have attempted a PhD without the support and encouragement of my parents Val and John. Without Tina, Lily and Fred to keep me focussed on the important things I would surely have given up before the final push. So this is it. Its finally finished. Now I can get on with something else…

“Your mind will answer most questions if you learn to relax and wait for the answer.”

– William Seward Burroughs II (1914 – 1997)

List of contents

List of figures

List of symbols

A	Area of exposed surface
$boundary_{t,\alpha}$	OLS CUSUM boundary at time, t and significance, α
CUSUM	Cumulative sum
cv_{α}	Critical value at significance, α
C_p	Specific heat capacity of air
DD_{τ}	Degree days to base τ
E	Measured energy consumption
$\hat{E}$	Predicted energy consumption
f_h	Heating energy consumption rate
f_{nh}	Non-heating energy consumption rate
H	Heat transfer coefficient
H_a	Portion of *H* due to air flow
H_f	Portion of *H* due to transmission through building fabric
N	Number of data points
OLS CUSUM	Cumulative sum of the ordinary least squares residuals
p	Number of model parameters
Q	Free gains
RMSE	Root mean squared error
SBC	Schwarz Bayesian criterion
SSE	Sum of the squared error
T_{in}	Thermostatic set point
T_{out}	Measured outside air temperature
U	Model residual
U	Transmission coefficient (U-value) of exposed surface
V	Volume flow rate of outdoor air entering the building
β	Heating coefficient
ε	Coefficient of variation of the RMSE
η	Efficiency
λ	Normalised event significance
ρ	Density of air
$\sigma_{BB,t}$	Standard deviation of a Brownian bridge at time t
τ	Change point / degree day base temperature

Chapter 1 Introduction

"Waste is worse than loss. The time is coming when every person who lays claim to ability will keep the question of waste before him constantly. The scope of thrift is limitless."

Thomas Alva Edison (1847 – 1931)

Energy management of buildings is an important function of many organisations such as local authorities. Controlling the financial and environmental costs associated with energy consumption helps meet greenhouse gas reduction targets and delivers the best value for public funds.

In a large portfolio of buildings, consumption patterns will both improve and deteriorate continuously. Faults occur and are rectified, maintenance is carried out and energy interventions are implemented. The activities within buildings change over time, new equipment is required, old equipment is upgraded or replaced. All these changes can affect energy consumption patterns.

For energy management purposes, there is benefit to be had from keeping these changes under close observation. A critical aspect of energy management is the collection and analysis of energy consumption data. Knowledge about historical changes in consumption patterns are the basis for assessing the impact of faults and the effectiveness of interventions. In general, these data aid in the identification and reduction of energy wastage.

Standard methodologies for the analysis of consumption data (see Chapter 2) were developed in the 1980's and are designed for use with the data available at the time, monthly billing data or manual meter readings. It is now common for organisations with large portfolios of buildings such as local authorities to systematically collect and store energy consumption data via automated meter reading networks. Such systems generate large databases of high-resolution (sub-daily) consumption data.

The large stores of historical consumption data generated by such systems contain valuable information about consumption patterns both in the past and present. A system operated by Leicester City Council, the main industrial sponsor of the present research, generated the dataset under analysis in this work (see Chapter 3).

Standard analytical methods summarise consumption patterns using a simple consumption model and identify any changes to those patterns over time. As part of the energy management process, changes are investigated and, if appropriate, corrective action can be taken. As they stand, these methods cannot be applied without the interpretation of an experienced analyst and thus cannot be easily applied on a large scale.

This research develops (see Chapter 4) and demonstrates (see Chapter 5) a methodology for automating event detection with respect to simple energy consumption models. An analysis is conducted on 738 half-hourly gas and electricity datasets, each covering a period of up to seven years.

Over 5,000 events where the pattern of energy consumption changed were identified and each event was characterised and quantified. Consistent periods of consumption between events are modelled in detail. This bulk statistical analysis of energy consumption data provides a detailed, objective, quantitative insight into changing consumption patterns across an entire building portfolio (see Chapter 6).

The analysis generates meta-data (data about data) which represents an event-oriented summary of consumption patterns for each individual gas or electricity dataset. Every change in consumption pattern is recorded and quantified. These results provide the basis for an approach to energy management that was previously not possible.

Previously the analyst would need to wade through the vast dataset manually looking for any interesting features. These meta-data can be used to direct the analyst towards the data of interest and to provide simplified contextual information for each event. Using the techniques presented in this work the analyst is free to spend their time interpreting and investigating events which are picked out of the data automatically.

1.1 Usage

In this work, the terms GHG management and energy management are used extensively. The term greenhouse gas management is defined as the systematic, ongoing process of reducing greenhouse gas emissions. Similarly, energy management is defined as the systematic, ongoing process of reducing energy wastage.

It should be noted that when the term GHG management is used it implies energy management since, for many organisations, energy consumption is a major source of carbon dioxide (CO_2) emissions and energy management forms a major element of greenhouse gas management.

Since this work is related to energy consumption, it will more often be the case that the term energy management will be used. When the term energy management is used, it implies the management of GHG emissions.

The most common quantity referred to in this work is energy. Many of the figures show energy consumption in one form or another. This work uses kiloWatt-hour (kWh) as the unit of energy rather than the SI equivalent (Joules) because they are the common unit used in industry.

The remainder of this chapter provides a more detailed introduction to this work. Section 1.2 describes the motivation for greenhouse gas management and the context within which local authorities are operating.

Section 1.3 describes a strategic approach to greenhouse gas management and identifies how analysis can be integrated into the process. Leicester City Council (LCC) is the major partner in this work. Section 1.4 provides a brief outline of the experience of LCC with setting up a city-wide automatic meter reading system.

The chapter ends with section 1.5 which specifies the aims and objectives of this work and section 1.6 which provides a brief chapter-by-chapter outline of the remainder of this document.

1.2 Context

The annual cost of energy consumption across UK local authorities runs into many millions of pounds. Leicester City Council (LCC) alone spends over £5 million a year on the energy and water requirements of its own premises (Leicester City Council 2011). This financial cost provides a significant incentive to carefully manage energy consumption and minimise wastage.

Many local authorities (including LCC) set up energy management systems in response to the energy crises of the 1970s. The main function of such energy management systems is to invest in energy saving measures with attractive financial payback. Since the 1970s, the influence on the global climate of greenhouse gases

(GHG) such as Carbon Dioxide (CO_2) has become increasingly apparent. Since the first reports of the Intergovernmental Panel on Climate Change (IPCC 1990a; IPCC 1990b; IPCC 1990c) the impact of energy consumption and wastage can no longer be seen as purely an economic consideration.

The primary drivers for energy conservation in the UK public sector are increasingly based on national legislation stemming from the United Nations Framework Convention on Climate Change (United Nations 1992) and the Kyoto Protocol (United Nations 1998). Being large public organisations, local authorities are under constant pressure to deliver best value for public funds. Local authorities are also influenced by national and local emission reduction targets. This makes them more likely to adopt a strategic approach to greenhouse gas management.

1.3 Strategic greenhouse gas management

Several sources of guidelines for developing greenhouse gas management strategies are available. This section provides a brief overview the basic approach.

The International Council for Local Environmental Initiatives (ICLEI) introduced the Cities for Climate Protection (CCP) campaign in 1993, it currently includes more than 800 local governments across six continents (ICLEI 2006). The CCP approach provides a core methodology for acting to reduce greenhouse gas emissions.

The CCP approach has been adopted by the UK Carbon Trust and developed into the Local Authority Carbon Management programme (LACM). The Carbon Trust approach to strategic carbon management is targeted towards the private sector (The Carbon Trust 2005), local authorities (The Carbon Trust 2006), higher education (The Carbon Trust 2006) and the National Health Service (The Carbon Trust 2008).

The Energy Star programme of the US Department of Energy (DOE) and Environmental Protection Agency (EPA) also offers guidelines for organisational energy management (USDOE 2004; Bennet and Whiting 2005). A schematic of the steps involved is shown in Figure 1.1.

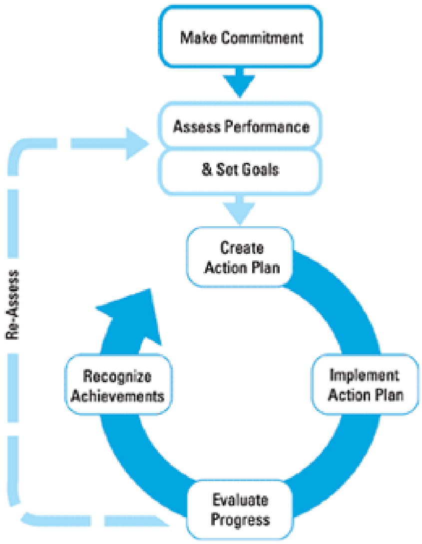


Figure 1.1: Steps of the Energy Star guidelines for energy management

The energy management matrix (BRECSU 1993; The Carbon Trust 2007) is a commonly used tool to assess organisational energy management. The matrix requires organisations to assess their performance in six key aspects of energy management. The categories include energy policy; organisation; motivation/training; information systems/performance measurement; marketing/communication and investment.

The CCP, LACM and Energy Star have many features in common and provide a framework for managing emissions. Much of this also maps to the content of the energy management matrix. The basic approach can be summarised as a series of interconnected processes which must be continually reviewed and updated.

1.3.1 Policy

The first step is to secure support from senior management. This includes the development of a policy document and the mobilisation of the organisation through setting up working teams and assigning responsibility to appropriate staff.

Support at the highest level is critical to implementing a strategy in a sustainable manner. There has been a clear series of initiatives driving this agenda forwards over the years (Energy Savings Trust 2011; European Commission 2011).

1.3.2 Strategy

Developing a strategy to deliver organisational emissions targets begins with the development of an inventory of emissions for the organisation. This provides a means to focus resources on activities where there is most potential for emissions reduction. The inventory must be kept up to date through regular review.

Armed with an emissions inventory it is possible to establish an emissions target relating to some baseline year (e.g. a 50% reduction on 1990 emission levels). A comparison of the latest emissions inventory with the emissions targets provides the headline figures with which to monitor the performance of the strategy.

Energy in buildings will very likely be a major element in a GHG reduction strategy. In terms of energy in buildings an emissions inventory may include a building-by-building list showing current emissions, historical emissions and possibly forecast emissions.

1.3.3 Action

The most important process which delivers GHG emission reductions is the development, prioritisation and implementation of energy efficiency interventions. Investment of resources should be allocated as efficiently as possible to ensure resources are not tied up unduly and to ensure the maximum impact on emissions. Financial instruments such as revolving funds can be used to balance lucrative energy efficiency investments with more expensive emissions reductions interventions.

The key working document is the action plan. This is basically a list of potential investments. These should be specific interventions with estimates of associated costs and emissions reductions. With very large building portfolios of several hundred buildings, the effectiveness of energy management can be largely reliant on the quality of the action plan.

A critical role of data analysis is to provide the diagnostic reports which help to populate the action plan with interventions. A rigorous statistical analysis has the potential to add a great deal of value to the energy management process by identifying and quantifying potential interventions automatically.

After implementation, each action should be evaluated against the predicted performance. A systematic process of evaluation ensures interventions are implemented effectively and enables improved estimation of savings for future interventions.

As a critical function of energy management, the data analysis developed in this work is intended to improve both the effectiveness and the efficiency of the process of generating and updating the action plan.

1.4 The experience of Leicester City Council

Leicester City Council (LCC) is a key partner in this research and has not only provided the data which will be used to demonstrate the methods developed herein but also helped to guide the research to ensure it delivered a useful output. Over the course of the project LCC were consulted regularly to gain an understanding of the practical problems faced when dealing with large numbers of detailed energy consumption datasets.

LCC was at the forefront of the move to high-resolution data collection over ten years ago. LCC has found that the additional information available through high-resolution data has been critical in identifying energy savings. Especially those due to weekend and overnight consumption that was previously ‘invisible’ being aggregated into monthly billing data. The detailed pattern of consumption across the day and week is now clearly visible. With simple visualisations of high-resolution data it is possible to clearly reveal energy wastage that would be obfuscated in monthly data. Information such as the timings of heating systems and the level of consumption during off-peak and unoccupied periods is readily available.

Once identified, wastage can often be acted upon very easily. For example, if a heating control system fails or is overridden then within a day or two the additional consumption caused by overnight heating would be apparent in energy consumption data being received automatically. Once the control is reconfigured then after one or two days the effect should be clear and after a week or two it is possible to establish reliable estimates about the avoided wastage. Also, if the control reconfiguration did not have the expected effect then it can be quickly resolved.

In LCC most of the ‘low-hanging fruit’ has been addressed by systematically reviewing each building in turn and identifying and acting upon wastage. As a result the building stock is performing better than ever and the data analysis task is more a continuous monitoring process. The key goal now is to identify the onset of wastage to act quickly to understand its cause and mitigate its effects.

LCC use Dynamat (Energy Metering Technology Ltd 2010) software to analyse their data. Like all such software the main focus is to provide access to charts of raw data, concentrating on the most recent data. In the main, historic data is not analysed beyond about one year. The software provides tools to generate simple performance indicators and diagnostics but doesn't provide a summary or league table of buildings showing relative performance.

With the increased information there is an equivalent increase in the amount of work required to analyse these data. At a large scale it is impractical to analyse data in this way. It takes time to generate appropriate charts. In most cases Analyst time is wasted confirming that buildings are still performing as they were in previous weeks. What is needed is a method to monitor these data automatically, to generate and compare energy performance indicators and to report any changes to performance in a timely fashion.

1.5 Aims and objectives

The aim of this work is to develop a method to model changing consumption patterns in buildings. To this end a method is required to identify events in half-hourly energy consumption data. The method should be automated because it will be used to analyse large numbers of building energy consumption datasets.

In order to achieve this aim, the following objectives were defined:

- To review energy management literature and assess the state of the art in energy consumption data analysis
- To identify appropriate models to describe energy consumption patterns
- To determine a statistical method capable of identifying changes in energy consumption patterns such as energy efficiency interventions and faults
- To prepare the data made available to this work for analysis
- To apply the chosen methodology to the 300 buildings for which data are available
- To demonstrate the value of this approach to energy management through detailed examples

The analysis is intended to provide information in support of municipal energy management efforts to reduce energy wastage and the corresponding CO_2 emissions. In particular, the analysis is focussed on detecting and quantifying changes to energy consumption patterns.

It should be noted at this point that the aim is not to develop a technique for energy managers to use directly. The techniques developed in this work are automated and can therefore be built into software systems.

The analysis conducted herein generates meta-data which describes the changing energy consumption patterns in each dataset under analysis. This event-oriented meta-data is intended to be the basis for further software which can use it to navigate and interpret the available datasets, directing the analyst towards the data of greatest interest.

1.6 Thesis summary

The thesis is presented in seven chapters and one appendix. The current chapter introduces the aims and objectives of the work and provides an introduction to the context in which the work is conducted.

Chapter 2 provides a detailed description of the common analytical approaches in energy management and introduces the theory behind the methodology implemented in this work. Empirical energy consumption models are described and the theory of structural change tests is introduced. Examples are provided using simulated data.

Chapter 3 describes the data collection and management system used to generate the data used in this work. The LCC data include over 700 individual datasets and cover up to seven years of consumption from over 300 buildings. Some summary statistics are presented and the steps taken to prepare these data for analysis are described.

Chapter 4 presents the methodology applied in this work. The development of a set of intelligent energy consumption models is described in detail. The application of a recursive event-detection algorithm is also described.

Chapter 5 presents the results of the automated analysis. Examples are provided to demonstrate the benefits of the approach on individual datasets. All available datasets are analysed and the resultant event-oriented meta-data are presented in summary.

Chapter 6 provides a discussion of how the results generated in this work can be utilised in municipal energy management systems. Examples are presented which indicate how the approach could be implemented as interactive software. The limitations of the models and event detection algorithm are also discussed.

Chapter 7 offers a summary of the key findings and conclusions arising from this work. The aims and objectives are revisited and discussed. The original contribution to knowledge arising from this research is summarised. Opportunities for future developments are suggested.

0 describes the problems identified with the raw data used in this work. The data were found to have a number of problems which needed to be resolved before the data could be usefully employed. A methodology is presented to identify and rectify these problem readings.

Chapter 2 Consumption models and event detection

"We shall not cease from exploration and the end of all our exploring will be to arrive where we started… and know the place for the first time"

– Thomas Stearns Eliot (1888 – 1965)

As described in Chapter 1, the quality and quantity of available energy consumption data have increased significantly in recent years. Standard approaches to data analysis in energy management were developed in the 1980's and are based on the analysis of monthly billing data or manual meter readings. Modern metering systems generate comparatively high resolution data and contain far more detail about consumption patterns. Standard models cannot capture this detail and standard analysis techniques cannot be automated. In order to take full advantage of this new resource, new data analysis techniques must be developed to automate the process and improve the resolution of results.

This chapter describes the existing data analysis techniques currently used within energy management. The benefits and limitations of existing techniques are highlighted. It also introduces statistical techniques which can be used to automate the detection of events in high-resolution energy consumption datasets.

Section 2.1 provides an introduction to energy monitoring and targeting (M&T) and energy measurement and verification (M&V). The main analytical methodologies of consumption modelling, event detection and event-oriented analysis are introduced.

Section 2.2 describes the standard CUSUM (CUmulative SUM) analysis technique used in energy M&T. The approach is one of event detection with respect to a consumption model. Advantages and limitations of the standard approach are highlighted.

Consumption models are discussed in section 2.2. In particular, the variable base degree day (VBDD) model which is used in the present work is introduced. A full derivation of the VBDD model is presented.

The key element of the standard CUSUM methodology is the event-detection stage. Section 2.5 provides a description of the statistical theory used in this work to automate event-detection. Statistical tests known as structural change tests are used to determine whether there is any evidence for events in the data.

2.1 Measurement and monitoring

Several motivations exist for the analysis of metered energy consumption data. It has been used to identify the effect of interventions (Lee 2000), to feed back information to energy consumers (Abrahamse, Steg et al. 2005), to demonstrate trends in appliance use (Firth, Lomas et al. 2008), to determine benchmarks for a given set of buildings (Hernandez, Burke et al. 2008) and to validate building energy simulations (Mahdavi, Orehounig et al. 2007).

The present work focuses on using energy consumption data to deliver information which can be used to highlight wasted energy, quantify the effects of changes in energy systems and otherwise support energy management activities. Of particular interest is the role of data analysis in large, multi-site organisations, where perhaps hundreds of datasets must be continuously monitored.

The information available through data analysis can have a major impact on consumption levels if appropriate action is taken. A 2004 study (The Carbon Trust 2007) found that small and medium-sized enterprises (SMEs) implemented an average annual saving of 5% of their carbon emissions (over £1,000 and 8.5 tCO_2 per site) through using advanced metering.

There is a large body of published guidance on analysis techniques specifically recommended for energy management. The majority of these publications fall into two distinct categories. In the UK ‘monitoring and targeting’ (M&T) is prevalent whilst ‘measurement and verification’ (M&V) was developed in the USA. The two traditions are distinct but have many similarities.

The basic principle in M&T is that continuous monitoring can help identify energy wastage (The Carbon Trust 2008). Consumption models are used to produce a projected expectation (or target) for consumption based on historical data. Measured consumption (usually through utility billing systems) is then compared to expectation. If measurement disagrees with expectation then investigation can be triggered and appropriate action can be taken.

The objective of M&V is to quantify savings achieved through investments in energy efficiency improvements (Efficiency Valuation Organisation 2007). Measurements taken before an intervention is implemented are used to construct a mathematical model of the consumption pattern. This is then used to forecast what consumption would have been without the intervention and from that, savings can be estimated.

So in general M&T is more exploratory; using consumption models and event-detection techniques to identify signs of wastage. M&V has a more limited scope and is more quantitative; using consumption projections to quantify the effects of known projects. The difference in emphasis reflects the difference in energy management culture in the two countries. In the UK the norm is for large organisations to have in-house energy management teams who will be responsible for delivering cost savings. In the USA outsourced energy service companies are more common and must prove they have delivered savings.

The current work draws on both traditions. The event-detection approach of M&T is taken as the basis but it is updated and used in conjunction with the consumption models and the event-oriented analysis from the M&V tradition. With this combined approach, the effect of arbitrary events where the pattern of energy consumption changes are automatically identified and quantified.

2.1.1 Monitoring and targeting

In the UK, organisations such as local authorities use M&T techniques to monitor energy consumption patterns in their own buildings (Vesma 2011). The basic M&T approach of consumption modelling and event-detection forms the core analytical framework for the present work.

In M&T energy consumption models such as the performance line (described in section 2.4.2 below) are used to establish expectations of consumption. When consumption differs significantly from expectation then an investigation can be triggered. The CUSUM (CUmulative SUM) technique (described in section 2.4.3 below) is used to monitor consumption and can identify even small changes to the prevailing pattern.

UK Government publications provide an account of the standard approach to M&T. Key guides include those published by the (then) Department for the Environment Transport and the Regions under the Energy Efficiency Best Practice Programme (BRECSU

1998; BRECSU 1998) and more recently by the Carbon Trust (The Carbon Trust 2005; The Carbon Trust 2006; The Carbon Trust 2008).

Other guides include those published by the Chartered Institution of Building Services Engineers (CIBSE 2006) and related guidance published under the US Federal Energy Management Programme (FEMP) (USDOE 2011) and Energy Star programme (USDOE 2011).

2.1.2 Measurement and verification

In the USA energy savings performance contracts (ESPC's) are widely used to attract third-party financing for energy saving projects (USDOE 2000). The assumption is that investment will result in energy cost savings. As such, accurately quantifying the impacts of interventions is important. This is the core task of M&V.

Organisations wishing to invest in energy efficiency use M&V techniques to verify savings, to allocate risks to the appropriate parties, to improve operations and maintenance and to reduce uncertainties. The primary reference which describes the M&V methodologies is the International Performance Measurement and Verification Protocol (Efficiency Valuation Organisation 2007).

M&V publications describe a series of energy consumption models which can be applied to monthly, weekly or daily measurements. Publications on M&V have more explicit links with professional and academic works than those connected with M&T. Most notable are the contributions of the American Society of Heating, Refrigeration and Air-Conditioning Engineers (ASHRAE).

The ASHRAE guideline 14-2002 for measurement of energy and demand savings (Haberl, Culp et al. 2002) provides guidelines on M&V techniques for Energy Service Companies (ESCO's) and their customers. (Claridge 1998) provides a historical perspective on M&V analysis techniques.

Since M&V is concerned with verifying savings, the subject of an analysis is always a known energy efficiency intervention. Energy savings cannot be measured directly. Instead, savings are determined by comparing energy use before and after implementation, making appropriate adjustments for changes in conditions (USDOE 2008).

The strength of M&V techniques is the suite of widely-used consumption models (see section 2.2.2) and the event-oriented techniques for working with these models as fitted to raw data before and after a known event. A projection of what consumption 'would have been' had the intervention not occurred is compared with measured consumption from after the intervention.

In the present work M&V techniques are applied to events which were not known to have occurred until they were detected in the raw data by statistical means.

2.2 Consumption models

Energy consumption models are used extensively in both M&V and M&T where they provide a means to represent systems which consume energy. They take the form of a mathematical description of the relationships between energy consumption and driving factors such as outside air temperature. Energy consumption models are at the core of all analysis methods described in this work

The models of interest to this research are empirical models for which measured energy consumption is the dependent variable and outside air temperature is the only independent variable. Such models have proven to be effective at predicting energy consumption in large numbers of buildings (Farouz, Baltazar-Cervantes et al. 2001; Haberl, Turner et al. 2002).

These models take the general form

$$\hat{E} = f(E, T_{out}) \qquad \textbf{2.1}$$

Where $\hat{E}$ represents energy consumption predicted by the model. The model is some function of measured energy consumption E and outside air temperature T_{out}. Model parameters are estimated for a given building or process when a model is fitted to measured consumption data.

Model parameters, along with the functional form of the model represent a simplified description of the consumption pattern. This enables the generation of consumption forecasts.

2.2.1 The Variable base degree-day model

The Variable base degree-day (VBDD) model (Fels 1986; Kissock, Haberl et al. 2003) is an example of a simple piece-wise regression against average outside air temperature. As its name suggests the VBDD relates consumption to degree days. Before discussing the degree day (degree days are described in section 2.4.1 below) this section will describe the properties of the VBDD model.

Mathematically, the VBDD model is defined as follows:

$$\widehat{\mathrm{E}} = \mathrm{f_{nh}} + \beta(\tau - T_{out})^{+} \qquad \mathbf{2.2}$$

Where $\hat{E}$ is the predicted rate of energy consumption, f_{nh} is the fixed, non-heating rate of consumption unrelated to weather, β is the heating coefficient, T_{out} represents the average outside air temperature and τ is the change point, the outside air temperature above which the heating coefficient has no effect. The superscript “+” indicates zero if the term in parentheses is negative. It will be shown (in section 2.4.1 below) that the term in parentheses is equal to the heating degree-days to base τ.

Parameter estimation

The VBDD model is an equation describing the nature of the expected relationship between fuel consumption and outside air temperature. When measured data are available it is possible to fit the VBDD model to these data by estimating the model parameters f_{nh}, β and τ. Parameter estimation is done using ordinary least squares (OLS) linear regression as shown in equations 2.3.

$$\beta = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \qquad \mathbf{2.3}$$

$$f_{nh} = \bar{y} - \beta\bar{x}$$

Where the ‘x’ elements are degree-days given by $(\tau - T_{out})^{+}$ and the ‘y’ values are measured consumption. Thus, the calculation is dependent on the change-point temperature. OLS regression requires that the change-point temperature is given in advance.

The VBDD model is fitted using a grid search algorithm. The ASHRAE inverse modelling toolkit (Haberl, Sreshthaputra et al. 2003; Kissock, Haberl et al. 2003) is software developed to fit VBDD and similar models to consumption data. It performs an algorithm whereby a range of balance point temperatures from 41°F to 80°F are all fitted to the model and the temperature which produces the highest coefficient of determination (r^2) is selected.

The model can be applied to consumption data at any resolution from daily to annual. Below daily resolution the model will suffer due to diurnal variation in internal gains, solar gains and due to the switching on and off of the heating system. The model works best at weekly resolution and above because there can also be interference due to weekly consumption patterns such as are caused by different occupancy on weekends. The VBDD model is applied in this work at daily resolution (see section 4.1.1) but adjustments are made to account for weekly variation (see section 4.1.2)

Model visualisation

The shape of the VBDD model is very distinctive. Visualising the model helps to clarify how its parameters relate to consumption patterns. Figure 2.1 shows the model with parameters set as f_{nh} = 50 kWh day^{-1}, β = -10 kWh ºC^{-1} day^{-1} and τ = 15.5ºC. In this example daily consumption is plotted against daily average outside air temperature.

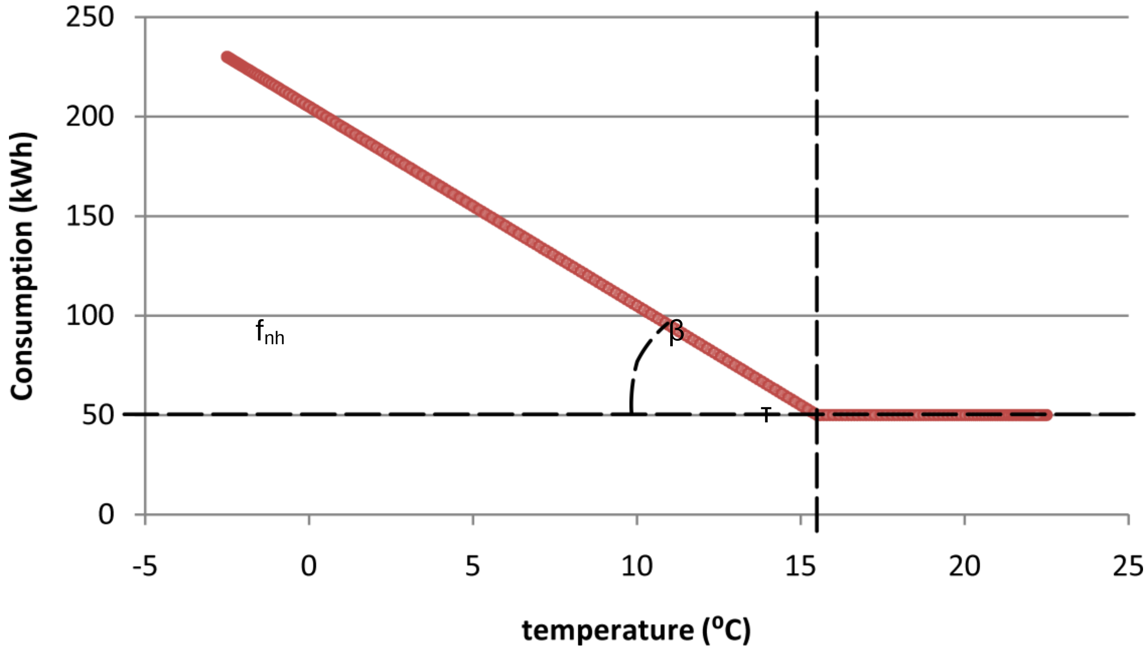


Figure 2.1: Basic VBDD model

The fixed consumption, f_{nh} is the minimum value which is reached only during warmer days. Consumption never falls below this value, even when no heating is required. The heating coefficient, β is the gradient of the line. The gradient only applies at temperatures below the change point, τ. The change point is the (outside) temperature above which heating is no longer needed.

Figure 2.2 shows the data from Figure 2.1 plotted as two time series. The data covers a whole year from summer to summer. Average daily outside air temperature is simulated as a simple sinusoidal pattern ranging from -3ºC to 23ºC with 365-day period. It can be seen in Figure 2.2 that during the summer, when average daily temperatures are above the balance point, daily consumption remains fixed. During the colder periods there is additional consumption which varies with temperature, this is commonly known as the heating season.

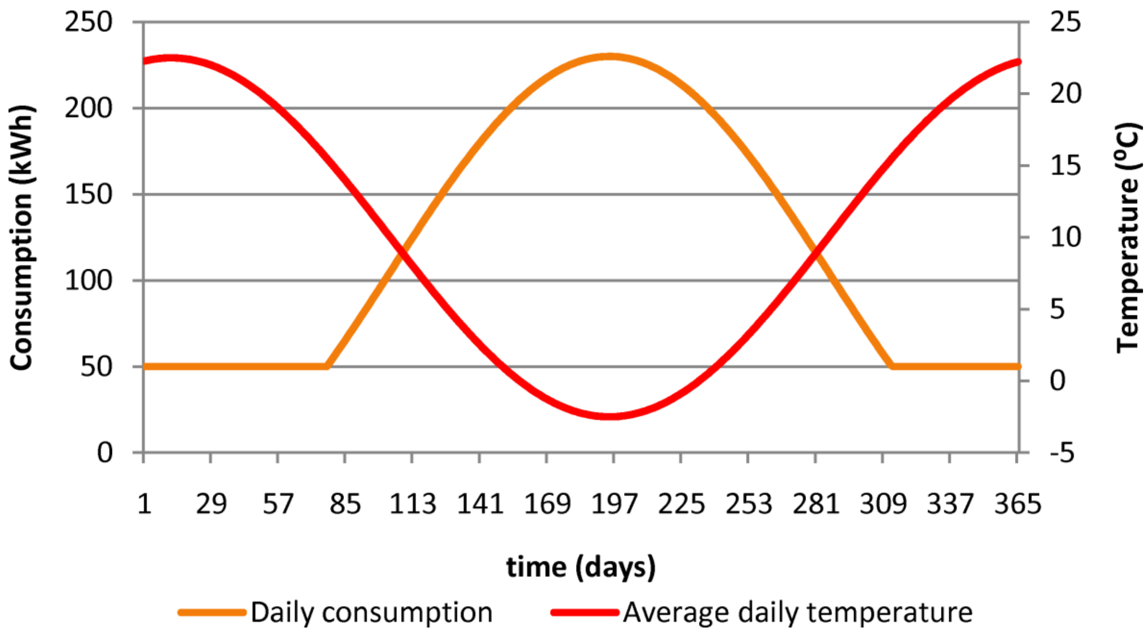


Figure 2.2: Consumption predicted by VBDD model given sinusoidal temperatures

In this purely abstract mathematical example, it is possible to inspect the effect of changing model parameters has on the pattern of consumption. This is helps clarify the physical meaning of each parameter. Figure 2.3 demonstrates the impact of changes to each parameter whilst keeping the temperature data the same.

Changes to the fixed component, f_{nh} has a fixed impact on every day of consumption. Estimating the value of f_{nh} for a building provides information regarding the base load which the building experiences regardless of any temperature related consumption. High values of f_{nh} indicate consumption unrelated to space-heating.

Changes to the heating coefficient, β only affect the heating season. Estimating the value of β for a building provides information regarding the thermal efficiency of the building envelope and heating system. High values can also indicate excessive infiltration.

Changes to the balance temperature, τ affect the length and intensity of the heating season. Estimating the value of τ for a building provides information regarding the thermal efficiency of the building and the thermostatic set points but it can also be affected by solar gain and occupancy.

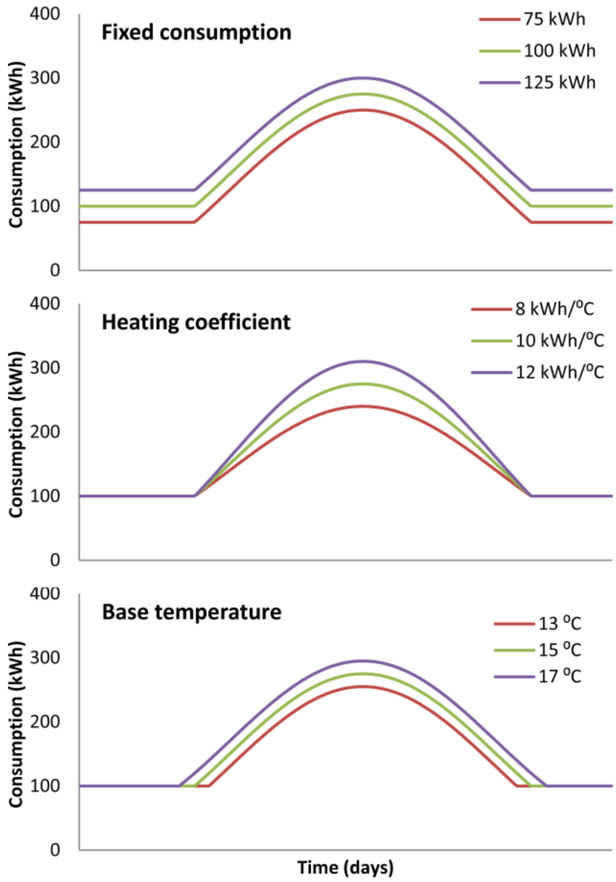


Figure 2.3: Changes to VBDD parameters

Derivation

It is accepted that buildings are very complex entities often with many separate and interdependent functions occurring within their walls. However, a simplified thermodynamic model predicts that energy consumption for space heating can be expected to closely correlate with outside air temperature.

The following theoretical treatment is a reproduction with minor modifications of that presented by Fels for the PRISM model (Fels 1986). The key assumptions are that free gains are fixed over a season, heat loss through evaporation is negligible as are thermal mass effects, and that enough energy is provided that the (fixed) thermostatic set point is always achieved.

Consider a building, with internal set point temperature, T_{in}. The rate of space heating energy required (E_h) to maintain this temperature can be determined by considering the rate of heat loss from the building. From Fourier's law of heat conduction it follows that the rate of heat loss is proportional to the temperature difference between inside (T_{in}) and outside (T_{out}).

$$E_h = H(T_{in} - T_{out}) \qquad (T_{out} < T_{in}) \tag{2.4}$$

Where H is the heat transfer coefficient of the building and has two contributions

$$H = H_a + H_f \tag{2.5}$$

Where H_a is due to air flow (infiltration and ventilation) and is given by

$$H_a = V\rho C_p \tag{2.6}$$

Where V is the volume flow rate of outdoor air entering the building, and ρ and C_p are the density and specific heat capacity of air respectively.

The contribution of transmission through the building fabric H_f is defined as

$$H_f = \sum U_j A_j \tag{2.7}$$

where U_j and A_j are the transmission coefficients and areas of each exposed surface of the building.

Some space heating energy is delivered by free gains Q (from appliances and lighting, occupants and the sun), the remainder is delivered by converting an energy source supplied at rate f_h, at an efficiency η.

$$E_h = \eta f_h + Q \tag{2.8}$$

Eliminating E_h from equations 2.4 and 2.8 gives:

$$f_h = \frac{H(T_{in} - T_{out}) - Q}{\eta} \qquad (T_{out} < T_{in}) \tag{2.9}$$

which may be rewritten as:

$$f_h = \beta(\tau - T_{out})^+ \tag{2.10}$$

Where the superscript positive sign indicates the bracketed expression is set to zero when it is negative and τ is a temperature below the thermostatic set point, T_{in}, by a value proportional to the free gains, Q and inversely proportional to the heat transfer coefficient, H

$$\tau = T_{in} - \frac{Q}{H} \qquad \textbf{2.11}$$

the additional supplied energy required to heat the space for every degree below the reference temperature is given by the heating coefficient, β

$$\beta = \frac{H}{\eta} \qquad \textbf{2.12}$$

If energy uses unrelated to space-heating such as hot water are assumed to be constant at a rate f_{nh}, then total rate of energy consumption, $\hat{E}$ is given by

$$\hat{E} = f_{nh} + f_h \qquad \textbf{2.13}$$

$$\hat{E} = f_{nh} + \beta(\tau - T_{out})^{+} \qquad \textbf{2.14}$$

Since heating energy is only needed when the direction of heat flow is from inside to outside. The bracketed expression is exactly equivalent to the definition of degree-days given in section 2.4.1 and comparison of equation 2.13 with equation 2.2 shows that the correct functional form has been derived with model parameters defined as follows

$$\beta = \frac{H}{\eta} = \frac{\left(V\rho C_p + \sum U_j A_j\right)}{\eta} \qquad \textbf{2.15}$$

$$\tau = T_{in} - \frac{Q}{H} = T_{in} - \frac{Q}{\left(V\rho C_p + \sum U_j A_j\right)} \qquad \textbf{2.16}$$

The above equations lead to several interesting and intuitively reasonable conclusions. Building characteristics will influence the model parameters in predictable ways. For example, an increase in internal gains Q can be expected to reduce the balance point temperature τ. An increase in efficiency η will reduce the value of β. An increase in infiltration rate, V or the overall u-value of the building envelope will increase both τ and β. Some limitations of this model are discussed in section 6.2.2.

2.2.2 Extensions to VBDD

The VBDD model has been developed into a collection of standard models based on a simple modified regression with outside air temperature only. (USDOE 2000; Haberl, Sreshthaputra et al. 2003; Kissock, Haberl et al. 2003; ASHRAE 2005). The models are recommended for M&V calculations of savings due to energy retrofits (Haberl, Culp et al. 2002).

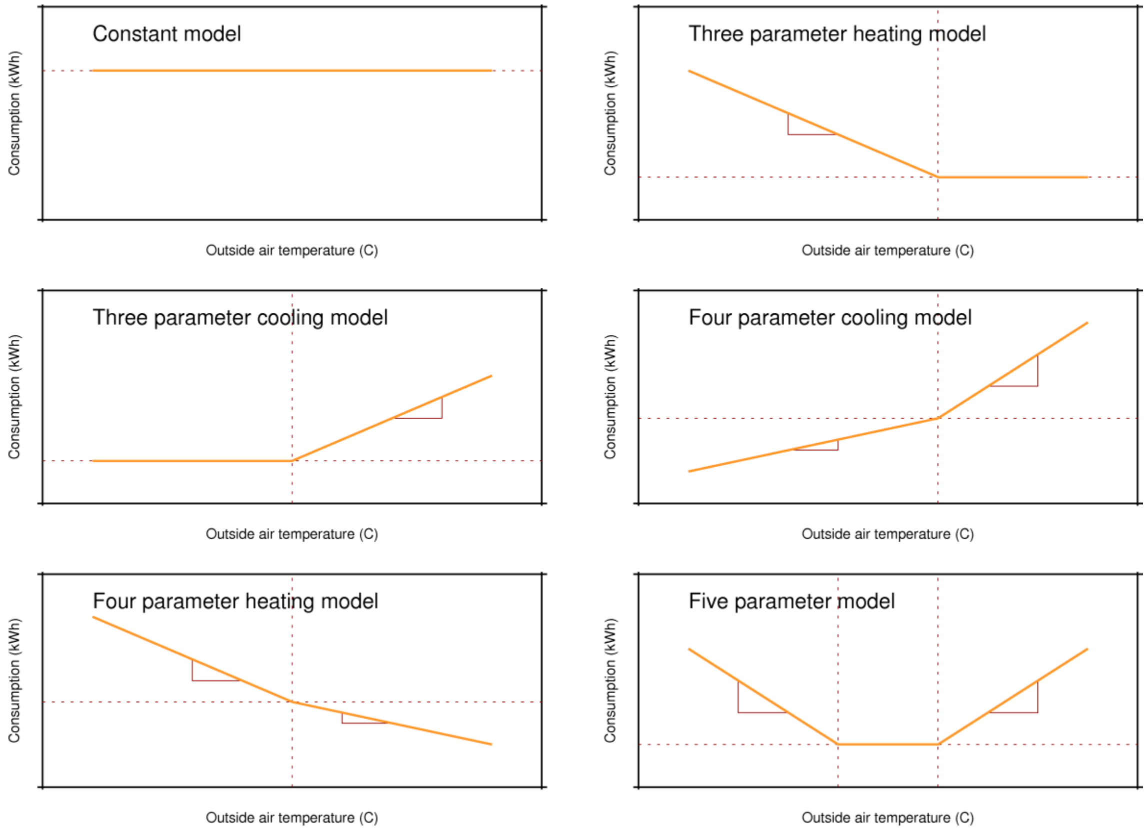


Figure 2.4: Variations of the VBDD model

These models are described in several publications and their shapes and names are shown in Figure 2.4. A comprehensive summary of alternative approaches is presented in (ASHRAE 2005). Since very few of the buildings under analysis in this work have active cooling systems the simple VBDD model (also known as the three parameter heating model) and the one-parameter constant model are suitable for use in this work.

2.2.3 Alternative modelling approaches

Aside from simple regression models several other approaches have been suggested for capturing energy consumption patterns. Most notable classes of models are various flavours of multiple regression (Ruch, Chen et al. 1993; Katipamula, Reddy et al. 1998; Kissock, Reddy et al. 1998; Kissock and Eger 2008), Fourier series (Dhar, Reddy et al. 1999) and artificial neural networks (ANN) (Kreider, Claridge et al. 1995; Kreider, Cohen et al. 1998). An assessment of these techniques concludes that they must be considered outside of the scope of this research.

The limited availability of independent variables precludes the use of multiple regression models. ANN modelling may be capable of producing a very accurate prediction of consumption but has the disadvantage that the model parameters generated are not easily associated with physical attributes.

Since the purpose is not to compare models but to develop a methodology for event-detection, alternative models such as those based on Fourier series are left as further work for comparison. For simplicity and to maintain a clear focus, only simple regression models are used in this work.

2.3 Event-oriented analysis

In the M&V tradition (Fels 1986; Claridge 1998; USDOE 2000; Haberl, Culp et al. 2002; Haberl, Sreshthaputra et al. 2003; Kissock, Haberl et al. 2003; ASHRAE 2005; Efficiency Valuation Organisation 2007) the primary goal is to quantify the effects of an energy saving intervention. Analysis is usually conducted with a known, dated intervention in mind.

In the present work the subject of analysis is not only energy saving interventions but other events such as faults and changes in consumption patterns in general which may be nothing to do with efforts to increase energy efficiency. An event may be the result of an intervention or any change to the underlying system, expected or unexpected. Events can cause consumption to increase or decrease but, by definition, they always cause consumption patterns to change.

In general the effects of an event can be determined by comparing the pattern of consumption before and after the event. The pattern of consumption can be determined by fitting any of the consumption models described above to the data in each period. In the examples below, variants of the VBDD model are applied to monthly data.

One method for estimating savings is to project the energy consumption pattern from the period before the event into the period after the event. This gives an estimate of what consumption ‘would have been’ had the event not happened at all. The method assumes that any change in the consumption pattern was wholly due to the event. Comparing this projection with measured consumption gives an estimate of the effect of the event.

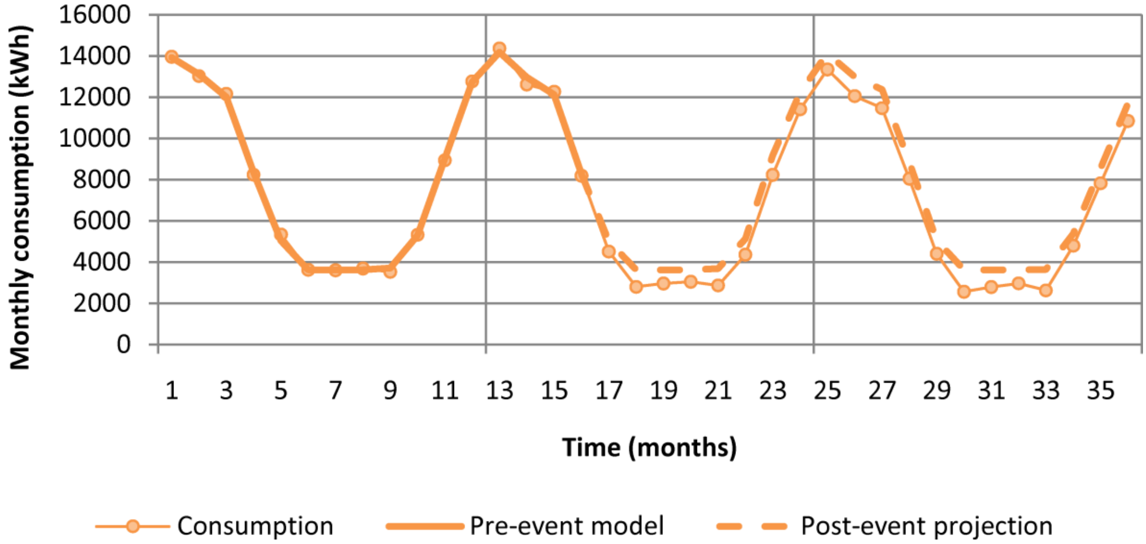


Figure 2.5: Savings estimation using projected consumption

A simple example is shown in Figure 2.5. In this simulated example an artificial event was introduced at month 16. The VBDD model is fitted to the data from months 1 – 16 and consumption is projected forwards through to month 36 using the estimated model parameters from the earlier period and temperature data from the later period.

Actual consumption is plotted as dots. The modelled consumption based on the period before the event is plotted as a thick line, the projection is shown as a dashed line whilst the pre-event model is a solid line.

To calculate the value of savings, simply sum the differences between the forecast consumption and measured consumption. The same approach can be used to calculate the effect of an event which caused extra consumption. This procedure is almost identical to the rebased CUSUM methodology illustrated in Figure 2.11.

When calculating energy savings in this way, ASHRAE's guideline 14-2002 stipulates that either the coefficient of variation of the RMSE of the baseline model should be less than 35% or the uncertainty of the savings estimate should be less than 50% at the 68% confidence level (Haberl, Culp et al. 2002).

An alternative approach is to calculate normalised annual consumption (NAC) for each modelled period. NAC is determined by calculating what consumption 'would have been' under one year of standard conditions. This may be historical average conditions or some base comparison year. NAC values have the advantage that they can be compared directly between periods and they provide a normalised savings figure which takes into account the impact of weather. However, since a period may be less than 12 months it may not cover the range of temperatures in the normalisation period and can thus lead to a problem of extrapolation.

2.4 The standard M&T approach

The standard approach to energy consumption data analysis published in M&T guides is to determine the pattern of consumption in a given building and to identify and possibly quantify any changes to that pattern over time. The present work develops this approach for use with large, high-resolution datasets.

This section describes the basic M&T approach as it is described in many publications. The fundamental concepts are degree days, the performance line and CUSUM analysis. Each will now be explained in turn.

2.4.1 Degree-days

To understand the analysis techniques which follow it is necessary to understand the notion of a degree-day. Heating degree-days are a measure of the severity and duration of cold weather (The Carbon Trust 2006). In a given period, colder weather will lead to a larger degree-day value for that period.

$$DD_\tau = (\tau - T_{out})^+ \qquad \textbf{2.17}$$

Where DD_τ represents degree days calculated to a base temperature τ. As mentioned above, the degree day definition is exactly equivalent to the bracketed expression in equation 2.2. This relationship is the basis for the performance line described in section 2.4.2 below.

Degree days are a summation over time of the amount by which the outside air temperature falls below a reference or ‘base’ temperature. Figure 2.6 shows temperature data for a three-day period. With the base temperature set to 15.5°C the degree-days are represented by the shaded area.

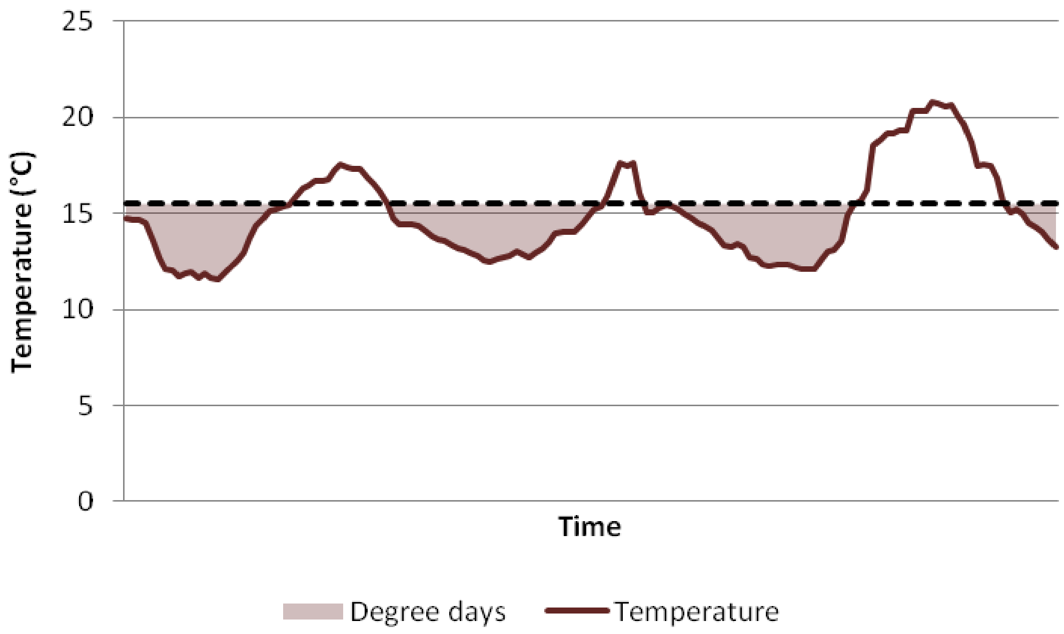


Figure 2.6: Degree-days

The degree day base temperature is sometimes referred to as the balance temperature. In the present work, ‘base temperature’ will be used when referring to degree days and the ‘change-point’ will be used when referring to the VBDD model. The degree-day base temperature can be interpreted as the outside air temperature below which heating is required. As equation 2.16 shows, the base temperature varies from building to building depending on the thermostatic set point temperature, available free gains, the thermal properties of the building envelope and the amount of air passing into the building from outside.

The base temperature is usually lower than the thermostatic set point because the building interior is heated by free gains such as from building occupants and solar gains. That is to say, if the outside temperature is 15.5°C then the internal temperature may reach the 18°C set-point even without the heating system being operational.

In the UK monthly degree-days are published for 18 different regions (see Figure 2.7) at standard base temperatures (15.5°C for most buildings, 18.5°C for hospitals). The implicit assumption here is that this is within the correct range for most UK buildings and that at the monthly resolution the impact of tweaking the degree-day base temperature is minimal.

Figure 2.7: UK Degree-day regions

Table 2.1 presents monthly degree day figures for the midlands (region six) for the year ending April 2010 (Vesma 2011). The 20-year average figures are also presented for each month. For example, it can be seen that October and November 2009 were particularly warm and December 2009 and January 2010 were particularly cold compared to the long term average in the midlands.

Table 2.1: Degree days for midland region (year ending April 2010)

Month	Degree days	20-year average degree days
May 2009	128	131
June 2009	73	62
July 2009	37	32
August 2009	30	32
September 2009	67	68
October 2009	135	160
November 2009	217	248
December 2009	395	343
January 20010	440	344
February 2010	348	304
March 2010	292	273
April 2010	205	206

2.4.2 The performance line

The performance line (Levermore 1992; BRECSU 1998; BRECSU 1998; The Carbon Trust 2006; The Carbon Trust 2008) is the most common energy consumption modelling method in energy M&T. This is because it can be easily implemented in a simple spreadsheet using freely available degree day data.

The performance line can be defined as a simple linear relationship between energy consumption and degree-days.

$$\hat{E} = \mathrm{f}_{\mathrm{nh}} + \beta DD_{15.5} \quad \textbf{2.18}$$

Where Ê is the predicted consumption, $DD_{15.5}$ is the degree-days calculated to base 15.5ºC, f_{nh} is the y-axis intercept of the performance line and β is its gradient. From equations 2.2, 2.17 and 2.18 the performance line can be seen to be equivalent to a reduced version of the VBDD model with the change-point temperature fixed at 15.5ºC (or 18.5 ºC).

For many buildings under normal operation, there is a high degree of correlation between degree-days and energy consumption for space heating. It is convenient to visualise the performance line on a scatter graph as in Figure 2.8.

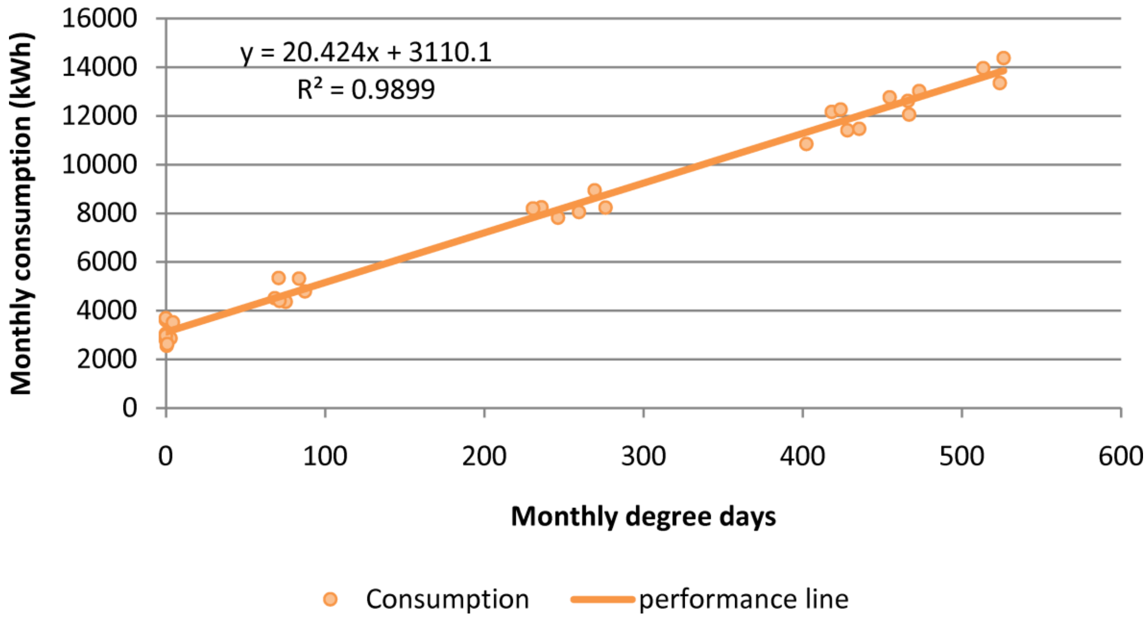


Figure 2.8: Performance line

Visualising this relationship can be used to determine whether space-heating is operating satisfactorily and diagnose problems (BRECSU 1996; BRECSU 1998). When the data are not linear it can be an indication of different kinds of problem (Harris 1989).

Being a simple linear model it is possible to estimate values for the gradient and y-intercept using ordinary least squares (OLS) regression with historical data. Figure 2.8 shows the estimated values for the linear parameters and a very high value for the coefficient of determination R^2. In section 2.4.3 below, this relatively small amount of scatter about the model will be shown to hide the effects of a significant event.

The performance line enables the direct disaggregation of space heating energy and energy not related to degree-days. The annual consumption which is independent of the weather can be calculated as the intercept, β multiplied by the number of data points in a year (e.g. 12 for monthly data).

The performance line can be used in conjunction with estimated future degree days to generate a prediction of consumption for energy budgeting purposes. It can also be used to check that energy consumption is continuing as normal e.g. by comparing a bill against the forecast value.

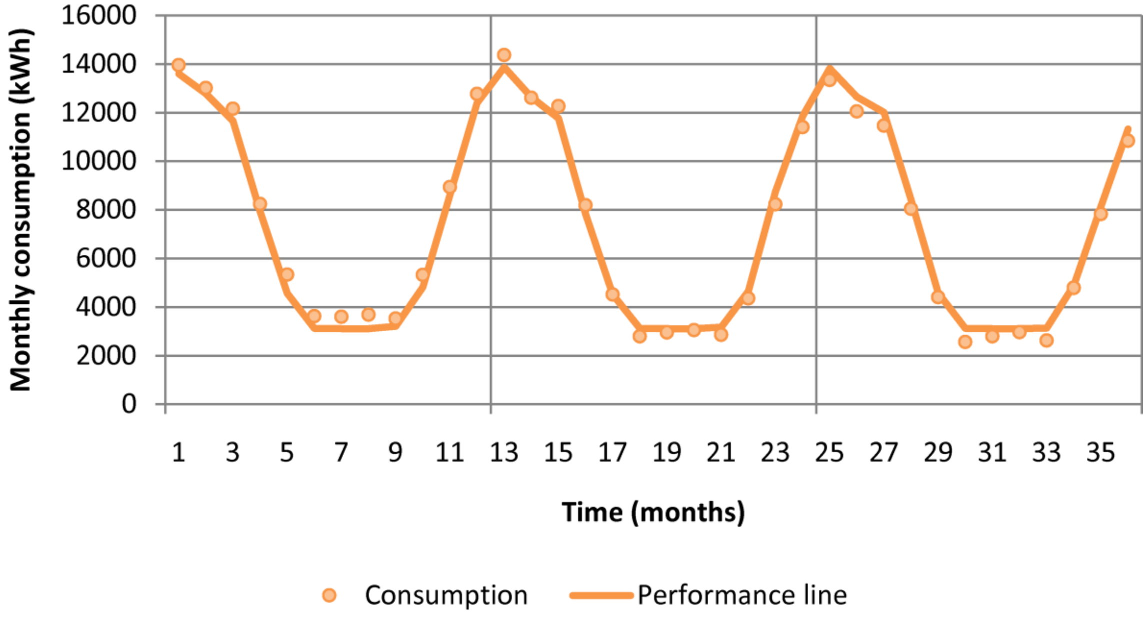


Figure 2.9: Consumption forecast using performance line

Figure 2.9 shows the same data as Figure 2.8 plotted against time. For each month there is a measured consumption value and a forecast consumption value using the known degree days for the month and the performance line parameters estimated from all available historic data using equation 2.18.

2.4.3 Cumulative sum analysis

The term CUSUM comes from a shortening of the phrase "**CU**mulative **SUM**". CUSUM is used extensively in energy management. Several of the guides already mentioned above describe a basic CUSUM methodology used in energy M&T (BRECSU 1998; BRECSU 1998; CIBSE 2006; The Carbon Trust 2006a; The Carbon Trust 2006b; The Carbon Trust 2008). This section describes this prevailing methodology.

In energy M&T it is common to use CUSUM analysis to monitor the consistency of consumption patterns over time as quantified by the performance line. A change in the parameters of the performance line indicates a change in building behaviour and an investigation can be triggered. Comparison of the performance line before and after a detected event can also be used to quantify the effect of an intervention or fault.

The methodology is deceptively simple, so much so that it has the potential to be misunderstood. None of the energy management publications offers any theoretical or statistical background to the technique. A review of the methodology for the analysis of electricity data from schools was published as part of this research (Stuart, Fleming et al. 2007).

The difference between measured and modelled consumption is referred to as the model residuals. In Figure 2.8 and Figure 2.9 the residuals are the distance between the measured points and the modelled consumption predicted by the performance line. More formally the residuals are defined as follows.

$$U_i = E_i - \hat{E}_i \tag{2.19}$$

Where the i^{th} model residual U_i is equal to the difference between the actual, measured consumption (E_i) and the predicted consumption ($\hat{E}_i$) for the i^{th} period.

The basis of CUSUM analysis is the construction and visual inspection of a CUSUM chart. The CUSUM statistic at a given point is calculated as the cumulative sum of the performance line residuals up to and including that point.

$$CUSUM_j = \sum_{i=1}^{j} U_i \tag{2.20}$$

Figure 2.10 shows the CUSUM chart resulting from the performance line presented in Figure 2.8. The accumulated residuals reveal that, for the first 16 months of data, consumption remained consistently above the performance line. After an abrupt change in month 17 the consumption pattern has shifted to below the performance line.

It is up to the analyst to interpret the CUSUM as a series of straight lines separated by kinks. The two periods of consistent consumption pattern are indicated by the two straight lines plotted over the CUSUM itself. It is here where there is most scope for misinterpretation and confusion. Section 2.4.4 describes the problems in detail.

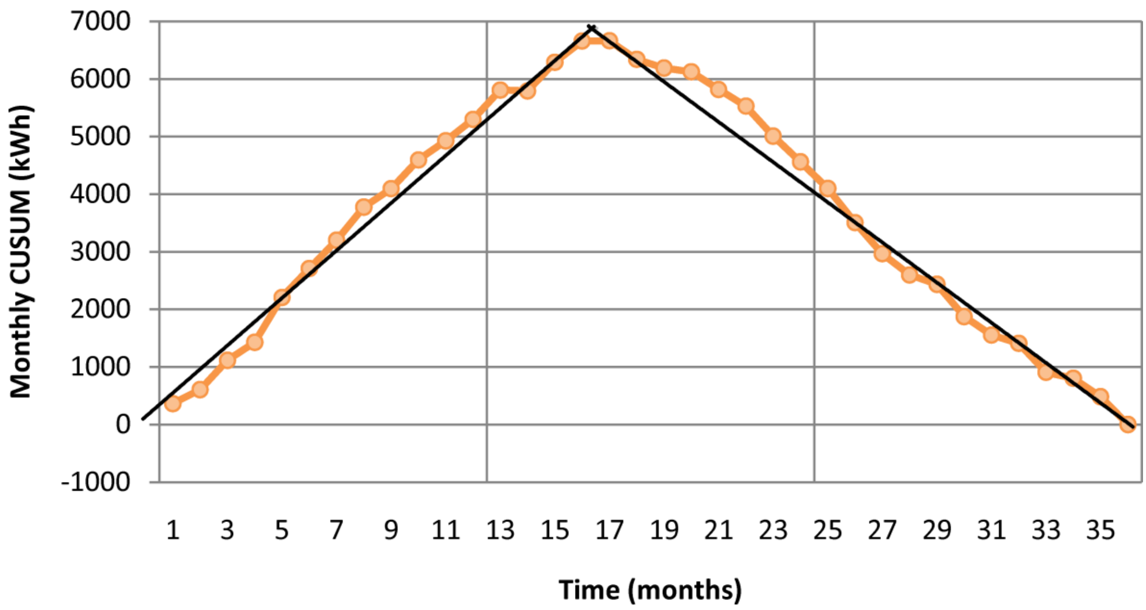


Figure 2.10: Basic CUSUM chart

Looking back to Figure 2.8 it is quite clear that the fit of the performance line was very good and there was no indication of any change in the performance line parameters when looking at the performance line alone. The CUSUM chart very clearly reveals an abrupt and persistent decrease in consumption after 16 months.

What appeared to be a small amount of scatter in the performance line is actually a small but persistent change in monthly consumption. During the first 16 months consumption is above the estimated performance line so the CUSUM accumulates positive residuals. At month 16 consumption falls below the estimated performance line the CUSUM begins to accumulate negative differences and the gradient shifts abruptly.

Looking closely at Figure 2.9 it is possible to identify this shift but it is small and difficult to see. CUSUM analysis emphasises these small fluctuations and allows for events which change the pattern of consumption to be identified and investigated.

CUSUM can also be used to quantify the effect of an event by ‘rebasing’ the model. This is done by fitting the performance line to the period before the event and forecasting forwards into the period after the event. The resulting CUSUM plot tracks the cumulative extra consumption or savings due to all events after the ‘training’ period.

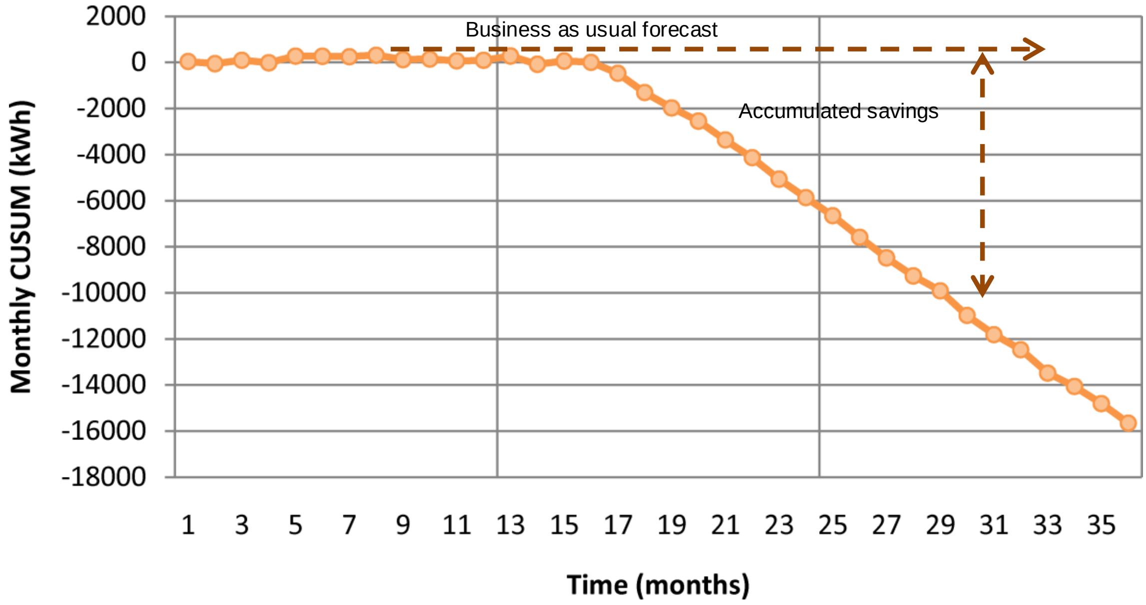


Figure 2.11: Rebased CUSUM

Figure 2.11 shows the rebased CUSUM for our example data. During the period before the event the rebased CUSUM shows a zero gradient. All things being equal, this path would be expected to continue with any accumulation of positive residuals being counteracted by negative residuals.

At month 17 observed consumption falls below the forecast consumption (based on data from before the event) and the CUSUM begins to drop. As the effects of the event are persistent there is an equivalent drop each month. By month 36 the figure shows about 16,000 kWh of savings have accumulated over 20 months. This method is equivalent to M&V savings calculations described in section 2.3.

This section has provided a description of the standard approach to the performance line and CUSUM in energy M&T. The approach is described in many of the guides referenced above with very little difference between guides. To differentiate it from OLS CUSUM described later in this chapter this analysis will hereafter be referred to as "standard CUSUM analysis".

2.4.4 Limitations

The combination of performance line and CUSUM analysis provides a systematic, analytical approach to monitoring which can be very effective in the hands of an experienced analyst. However, there are problems with the approach which means it is difficult to apply on a large scale over hundreds of high-resolution datasets.

The performance line is reliant on degree-days which need to be calculated to a given base temperature. For buildings in the UK the degree-day base temperature is usually assumed to be 15.5ºC. The base temperature is actually dependent on the building thermal performance and internal gains. Day (Day, Knight et al. 2003) suggest the use of the wrong degree-day base temperature can lead to misinterpretation of the performance line.

Day also suggests a method by which the degree-day base temperature can be determined with monthly datasets. However, in the present work the availability of high resolution data allows for the simple estimation of the degree-day base temperature using the VBDD model (see section 2.2.1).

One problem with CUSUM analysis is that the analyst is responsible for visually interpreting the fluctuations revealed in the CUSUM chart. The analysis needs a human in the loop and this constitutes a significant bottle-neck in the process. This also creates a problem with consistency of interpretation. The technique is heavily reliant on the interpretation of the analyst. The same dataset can be interpreted differently using the same technique. Significant experience is required to achieve consistency of interpretation.

The charts in Figure 2.12 were generated by simulation to demonstrate the problem. Four CUSUM charts are shown, each with the same event but with increasing amounts of random noise from top-left, through top-right to bottom-left and bottom-right. In this case the data are simulated and the interpretation shown in Figure 2.10 correctly represents the changes to the underlying model.

As more noise is added the ‘real’ event disappears and the CUSUM charts pick out the ebb and flow of random variations. There is no objective means in the standard methodology to determine whether these random fluctuations should be identified as real events or ignored.

The analyst must apply their experience with the analysis and their knowledge of the site to determine an appropriate interpretation. It is easy to see how the datasets may be interpreted as indicated on the charts with events being identified falsely.

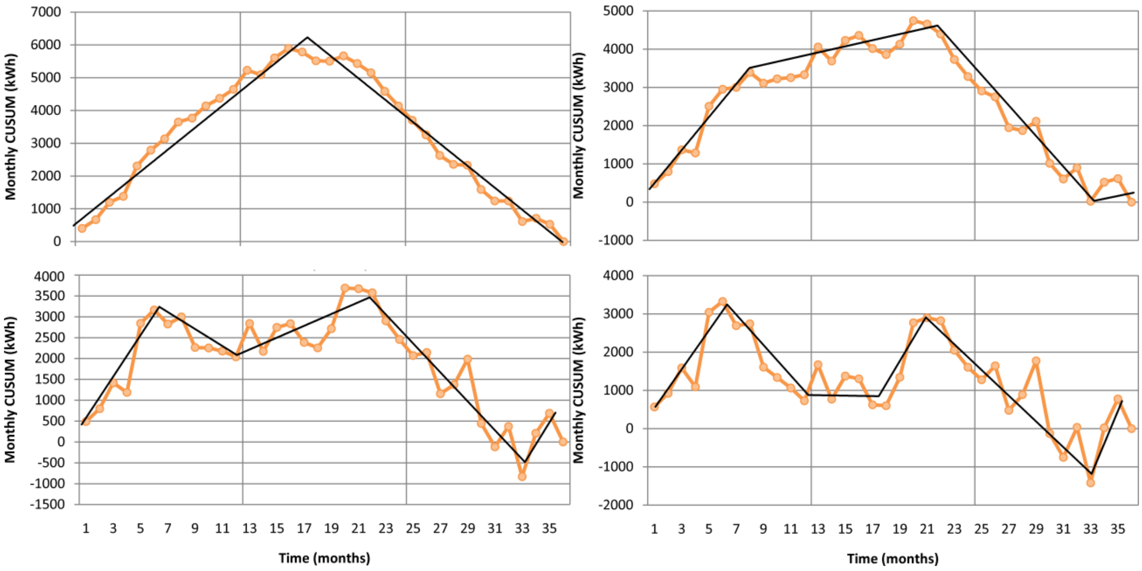


Figure 2.12: CUSUM charts showing one event and increasing background noise

An automated method to identify events would not only enable a quick analysis of multiple datasets but would also improve the consistency of event detection. An objective, quantitative means of selecting events has the potential to replace the use of the standard technique.

2.5 Event detection

Energy consumption models described in section 2.2 capture the relationship between environmental variables and energy consumption. When these models are applied to a set of measured data there is an implicit assumption that this relationship remains unchanged over the period under analysis. Event-detection methodologies such as CUSUM discussed in section 2.4.3 are employed to investigate this assumption.

Automatically detecting events is the main aim of this work. An automated technique would provide a new kind of event-oriented information to energy management. Existing techniques described above can identify if the consumption pattern changes but it is not possible to do this automatically and consistently.

Very few studies suggest means to interrogate changes in energy consumption patterns. Cluster analysis has been used to classify each day of consumption by the shape and magnitude of the profile (van Wijk and van Selow 1999). To some extent the issue is addressed (Fels 1986) by looking at a rolling assessment of Normalised

Annual Consumption (NAC, see section 2.3 for details). These methods are not designed to actively search for events.

No peer-reviewed studies have been found for which identifying changes in energy consumption patterns is the primary objective. This is the primary objective of the present work. In this section the CUSUM methodology as described in section 2.4.3 is explored in more detail and an automated approach is introduced.

2.5.1 The origins of cumulative sum analysis

Cumulative sum charts were originally developed as an alternative to control charts as proposed by (Shewhart 1938). The most common aim of such analyses is to detect changes in the quality of output from a continuous production process. The analyses are designed to detect changes in some measured parameter over time.

Control charts (or Shewhart charts) typically compare each observation with a fixed target value. If the observation crosses predefined action limits either side of that value then action is required. Action in this context may be to halt the process and reset a machine.

Assuming a normally distributed random fluctuation within certain tolerances is acceptable and should be ignored, action limits can be set relative to the measured mean (μ) and standard deviation (σ). For example, limits set at $\mu \mp 3.09\sigma$ to give a significance of 0.002 would give a one in five hundred chance of falsely detecting a change in the process mean (and e.g. stopping the machine unnecessarily).

Some control charts also include warning limits. These may be set at e.g. $\mu \mp 1.96\sigma$ to give a significance of 0.05 or a one in twenty chance of any one measurement crossing the boundary. Action is taken if a given number of recent observations fall between the warning limits and the action limits.

The CUSUM procedure (Page 1954) represents a fundamental change to the control chart methodologies that preceded it. Rather than simply comparing each observation with a reference or target value, the CUSUM method accumulates differences from the target.

$$CUSUM_j = \sum_{i=1}^{j} (x_i - k) \qquad \textbf{2.21}$$

Where x_i is the i^{th} observation and k is a fixed target value. The CUSUM statistic is the cumulative sum of the differences from the target value. Each observation is given a score (deviation from target) and scores are accumulated across all observations.

Under a one-sided CUSUM scheme (only looking for high values for example) action is taken if the current sample point raises the CUSUM statistic more than a stated amount, h, above the previous lowest point of the sample path. Such schemes are often 'tuned' to the problem at hand so that the process is controlled within appropriate limits (Page 1961).

2.5.2 Structural change tests

CUSUM is part of the wider category of statistical tests known as structural change tests or tests for parameter instability (Zeileis 2003). Such tests are designed to automatically detect events in time series data. From (Zeileis, Kleiber et al. 2003):

> *"Structural change is of central interest in many fields of research and data analysis: to learn if, when and how the structure of the data generating mechanism underlying a set of observations changes"*

The quote concisely reflects the three main stages involved in this kind of analysis: 'if'; 'when'; and 'how'. The first stage ('if') is to identify whether there is any evidence for a change at all. Where evidence is found then the most likely event can be pinpointed ('when'). Finally, with identified events, a comparison between data from before and after the event can reveal the nature of the change ('how').

In the case of energy consumption observations the underlying data-generating mechanism is the wider building energy system including the building envelope, equipment, occupants and operators. Detecting changes in this system is of fundamental importance to energy management.

2.5.3 Evidence for change

The key element of this analysis is a statistical test to assess whether there is any evidence for changes in model parameters over time. The presence of an event in a dataset can be determined by testing the following hypothesis:

H_0: There is no event in the dataset, model parameters remain constant

H_a: There is at least one event in the dataset

The core test methodology begins with fitting a model to the data and analysing the fluctuations in model residuals. The null hypothesis (H_0) of no change is rejected if the fluctuations are improbably large.

OLS CUSUM

Several transformations designed to capture fluctuations in model residuals have been suggested including variants of the CUSUM or MOSUM (**Mo**ving **sum**). Since it is close to existing M&T methods and is easy to calculate, the CUSUM of ordinary least squared (OLS) residuals or OLS CUSUM is used in this work. The OLS CUSUM (Ploberger and Krämer 1992) is very similar to the standard CUSUM analysis described in section 2.4.3.

The CUSUM statistic at time, t (between 0 and 1) is defined as follows.

$$\mathrm{OLS\ CUSUM_t} = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor}\mathrm{U}_i \qquad \textbf{2.22}$$

The OLS CUSUM is normalised by the standard deviation of the residuals (σ) and the square root of the number of data points (n). This normalisation is not used in the standard M&T method but is a critical factor since it allows for the magnitude of variations to be meaningfully interpreted. In the case of OLS CUSUM the residuals, U, are always determined by fitting a model using ordinary least squares so these variations are the cumulative divergence from the model.

By definition the CUSUM statistic begins at zero at time t=0. Due to the model fitting process the total of the model residuals will always be zero. It follows that the final value of the CUSUM (at time, t=1), is also zero. The OLS CUSUM method can be applied using any model and data at any resolution, it is based on OLS model residuals so, as long as the model is appropriate, any resolution data can be analysed in this way.

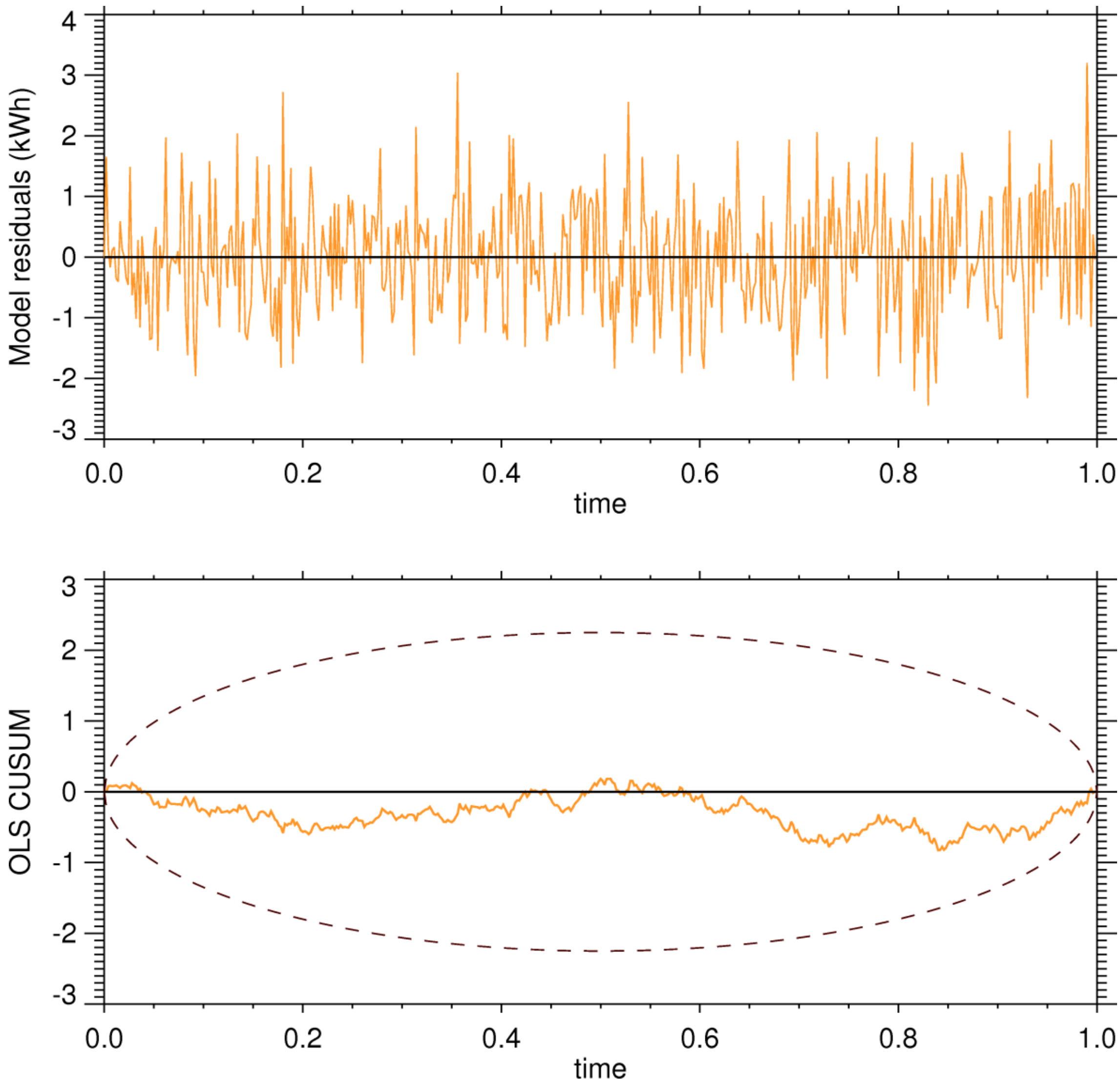


Figure 2.13: OLS CUSUM with 0.1% boundaries for random, normal data

Crucially, under the null hypothesis (of no events) the model residuals are expected to be a normally distributed random dataset with zero mean and standard deviation equal to one. Figure 2.13 shows a simulated dataset with these properties in the upper panel. The resultant OLS CUSUM is shown in the lower panel.

The precise shape of the CUSUM chart depends on the variation in the mean of the residuals. That is, although the residuals have a mean of zero, variations in the local mean along the dataset will cause the CUSUM to rise and fall. Any movement away from zero is due to the heterogeneous structure of the model residuals.

The Brownian bridge

To interpret the statistical inferences which follow it is necessary to take a minor diversion to discuss the statistical properties of random sequences. In particular, special cases known as the Wiener process and the Brownian bridge (Karatzas and Shreve 1991). The Weiner process is essentially a random walk, a sequence whereby each successive value is a random (normally distributed) distance from the last. It is possible to construct a Wiener process by accumulating a normally distributed, random variable.

A Brownian bridge is similar to the Weiner process but tied to zero at both ends. It is possible to construct a Brownian bridge by accumulating a normally distributed random variable with a mean equal to zero. This can easily be achieved by calculating the mean of the random dataset and subtracting it from each point before they are accumulated.

The Brownian bridge is of interest in this work because, under the null hypothesis, the model residuals will be normally distributed and homogeneous across the length of the data with a mean of zero. Therefore, under the null hypothesis, the OLS CUSUM conforms to the definition of a Brownian bridge.

Conversely, if a given OLS CUSUM is a Brownian bridge then the model residuals are homogeneous and there is no evidence with which to reject the null hypothesis. If the OLS CUSUM is not a Brownian bridge then the null hypothesis must be rejected and there is clear evidence for structural change.

In order to establish whether the null hypothesis should be rejected it is necessary to establish if the divergence from zero is improbably large. A statistical test is required to reject the null hypothesis that the OLS CUSUM is in fact a Brownian bridge.

The properties of a Brownian bridge are well understood; the largest absolute value in a Brownian bridge follows the Kolmogorov-Smirnov distribution. The critical values of which are shown at various significance levels in Table 2.2.

For example there is a 0.1% chance that the maximum value of a true Brownian bridge will exceed 1.949. Thus, if we impose a 0.1% (0.001) limit on the significance then the null hypothesis is rejected if an OLS CUSUM exceeds an absolute value of 1.949. The extreme value is considered so unlikely that it must be due to a real artefact in the data.

Table 2.2: Critical values for a constant boundary at a range of significance levels

Significance, α	Critical value
0.1 (10%)	1.224
0.05 (5%)	1.358
0.01 (1%)	1.628
0.005 (0.5%)	1.731
0.001 (0.1%)	1.949

However, the critical values above are only useful for identifying whether a given dataset conforms to the characteristics of a Brownian bridge. Zeileis argues (Zeileis 2004) that the standard boundary given by the Kolmogorov distribution is more sensitive to changes in the middle of the dataset and relatively weak at the beginning and the end. He further suggests that it is desirable to develop a boundary which varies with t in proportional to the standard deviation of the Brownian bridge.

The benefit of such a boundary over the standard boundary are that it is as equally sensitive to changes early and late in the dataset as it is to changes in the middle. The standard deviation of a Brownian bridge (σ_{BB}) varies with t according to the following relationship.

$$\sigma_{BB,t} = \sqrt{t \times (1 - t)} \qquad \textbf{2.23}$$

This gives the curved shape to the boundary shown in the bottom panel of Figure 2.13. At either end of the scale (t = 0 and t = 1) the standard deviation is zero because the value of the Brownian bridge is known to be zero. In the middle (t = 0.5) the standard deviation is maximised at 0.5. Table 2.3 shows critical values at various levels of significance reproduced from (Zeileis 2004; Zeileis 2008). These critical values show the magnitude of the curved boundary at each level of significance.

Table 2.3: Critical values for an alternative boundary at a range of significance levels

Significance, α	Critical value
0.1 (10%)	3.133
0.05 (5%)	3.375
0.01 (1%)	3.833
0.005 (0.5%)	4.000
0.001 (0.1%)	4.500

To determine the value of the boundary at any time, t for a given level of significance, simply multiply the standard deviation of the Brownian bridge ($\sigma_{BB,t}$) by the appropriate critical value from Table 2.3 (cv_α).

$$boundary_{t,\alpha} = cv_\alpha \times \sigma_{BB,t} \tag{2.24}$$

Figure 2.14 demonstrates the use of this boundary at 10%, 1.0% and 0.1% significance with one thousand randomly generated Brownian bridges. Each dataset represents an individual case where the null hypothesis holds true and the OLS CUSUM should not cross the boundary. The 1,000 datasets are plotted over one another and mostly overlap in the centre of the figure around zero. Due to the nature of the Brownian bridge, the datasets diverge from zero according to the shape described in equation 2.23 with a maximum divergence at around t = 0.5.

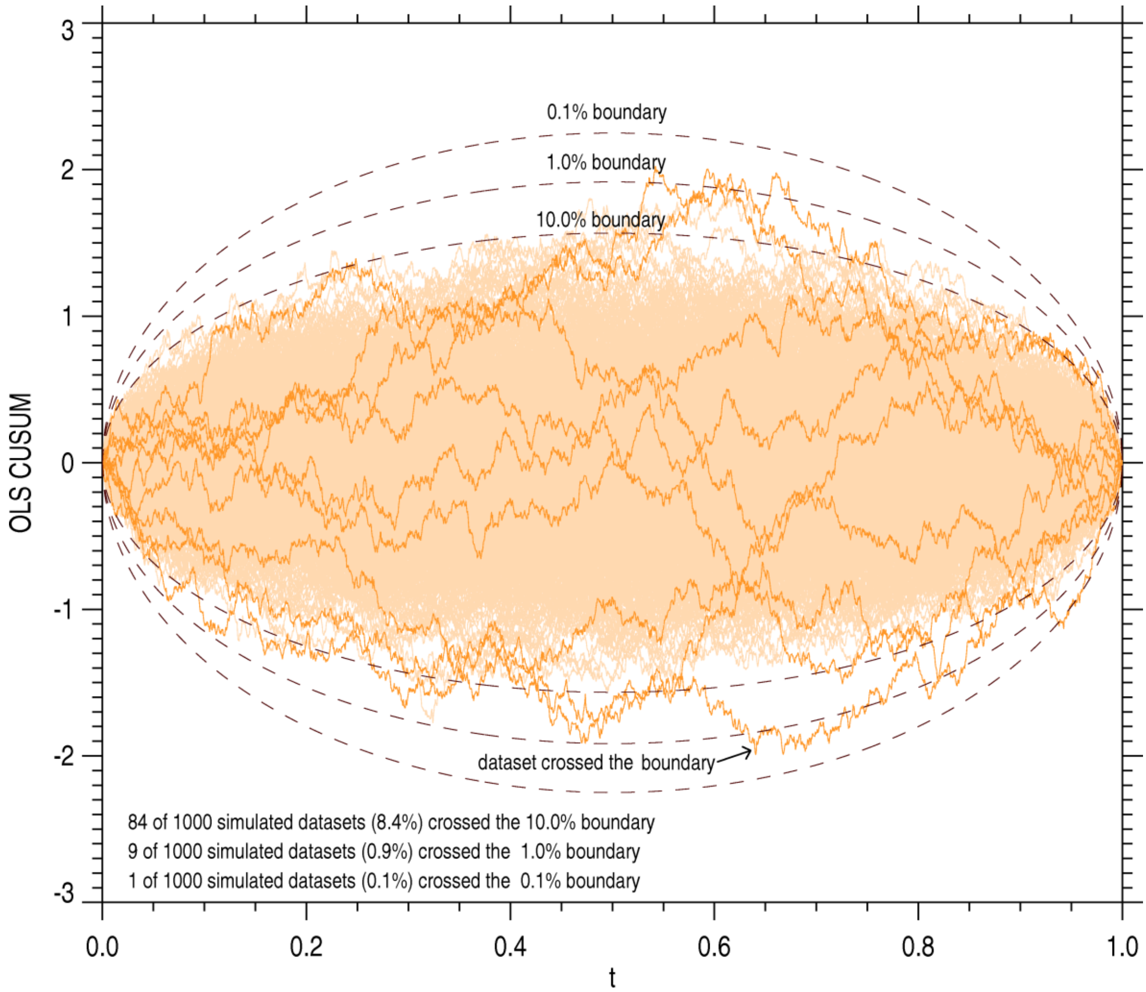


Figure 2.14: One thousand true Brownian bridges (9 cross the 1.0% boundary)

The figure also shows the boundaries at 10%, 1.0% and 0.1% significance. Of the 1,000 randomly generated Brownian bridges, 84 (8.4%) crossed the 10% boundary, 9 (0.9%) crossed the 1.0% boundary (these are highlighted) and 1 (0.1%) crossed the 0.1% boundary. This neatly demonstrates how the boundaries should be interpreted; there is a 1% chance that a true Brownian bridge will cross the 1% boundary. At 1.0% significance it is expected that the null hypothesis will be rejected for around 1.0% of true Brownian bridges.

2.5.4 Dating changes

Once the null hypothesis of no change is rejected it is necessary to pinpoint the event which caused the divergence. The assumption herein, which may or may not be the case, is that the change occurred at a distinct point in time or over such a short period that it can be approximated to a point in time (i.e. only a few time steps).

Figure 2.15 shows a simulated dataset where the null hypothesis is rejected. It is comparable with that in Figure 2.13 but a small step change has been introduced. The OLS CUSUM is calculated with respect to a constant, one parameter model. The change causes the OLS CUSUM to accumulate positive values before the event which are exactly cancelled out by accumulated negative values after the event.

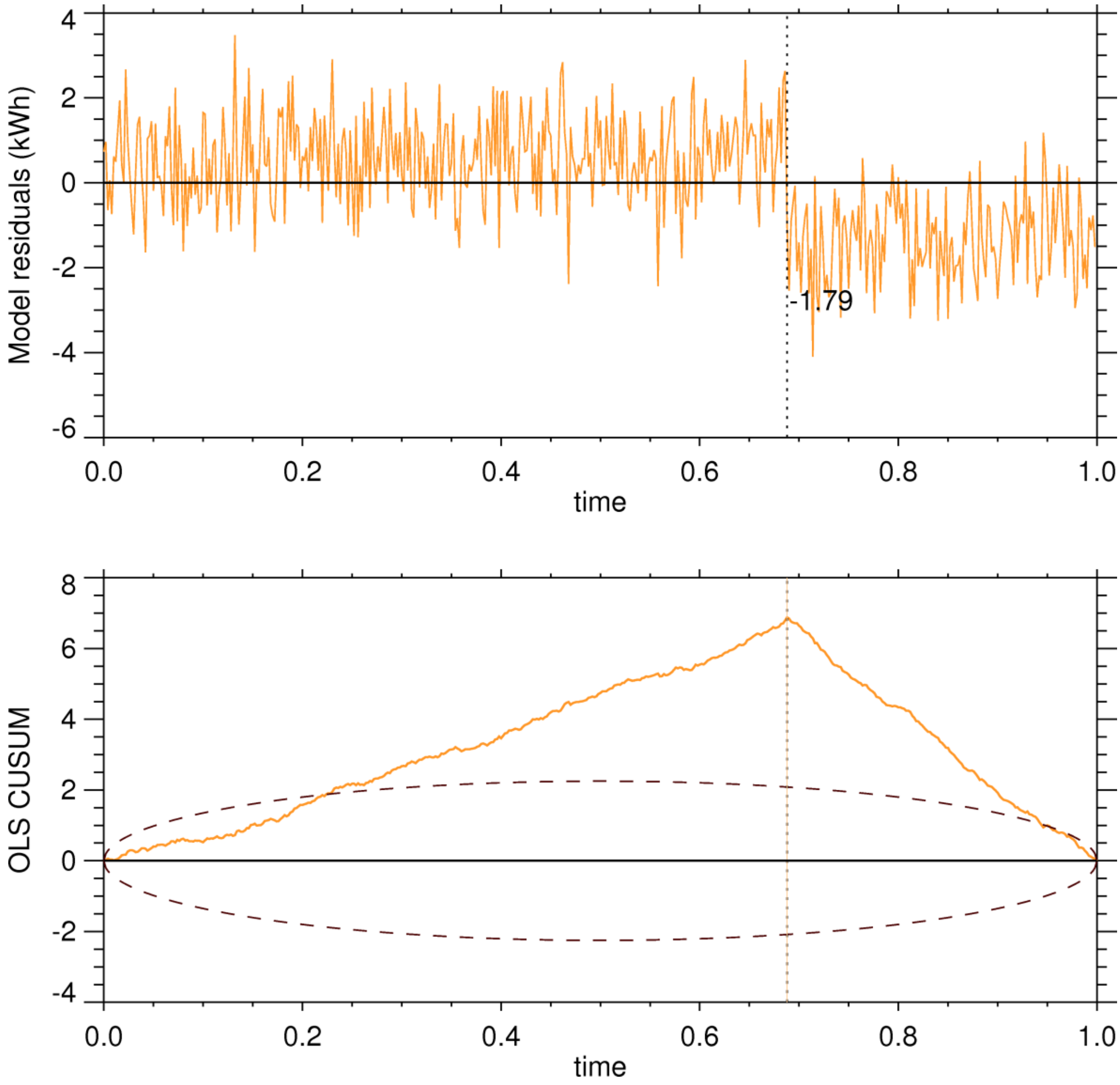


Figure 2.15: OLS CUSUM with boundaries for data with a single simulated event

The null hypothesis is clearly rejected as the 0.5% significance boundary is crossed. The next step is to pinpoint the most likely location for the onset of the event. Zeileis argues that the maximum absolute OLS CUSUM value is a good candidate. An alternative is to identify the point which exceeds the boundary by the greatest amount.

The difference in mean residual is maximised on either side of the maximum absolute CUSUM. However, the sensitivity is equally distributed across the dataset at the point which is furthest beyond the boundary. In this particular dataset these points are the same and match the location of the simulated event.

The OLS CUSUM method described above provides a general, automated means of detecting changes in modelled time-series data. When applied to energy consumption data with respect to consumption models described in section 2.2 this represents a method for detecting changes in consumption patterns.

This method is developed and applied in the present work; Chapter 4 describes the development of the models and application of OLS CUSUM event detection; Chapter 5 describes the results of the application of this method to consumption data provided by Leicester City Council. Before these, Chapter 3 will introduce the raw data to be analysed.

Chapter 3 Data collection and management

"Those who cannot remember the past are condemned to repeat it."

– George Santayana (1863 – 1952)

Before data can be analysed they must be generated, collated and stored. Energy management information systems combine systems for data collection, management and analysis. The chapter begins with section 3.1 which describes the practicalities of data collection and management.

The data under analysis in this work have been collected from over 300 local authority buildings in the city of Leicester. Section 3.2 presents an overview of the system which generated these data and a description of their format.

Managing large datasets is a complex task. Interrogating hundreds of datasets, each containing several thousand data points requires a highly systematic approach. The tools used to access and manipulate these data are described in section 3.3.

3.1 Energy management information systems

Systems for the collection, analysis and reporting of data relating to energy performance are collectively known as energy management information systems (EMIS). EMIS provide the information which underpins the energy management process (BRECSU 1998).

A detailed description of the component parts of an EMIS is given in (Hooke, Landry et al. 2003). In general, an EMIS comprises hardware, software and management systems which combine to deliver energy management information.

Hardware such as sensors and instruments (e.g. energy metering), communications infrastructure and data storage are necessary to generate the data, transfer it to a central location and to store the data for later retrieval. Software provides data management and analysis tools for the underlying communications, storage and reporting of data.

3.1.1 Data sources, resolution and accuracy

The quality of information it is possible to extract from energy consumption data depends largely on their accuracy and resolution. Both of these aspects of available consumption data have improved enormously with the introduction of automated meter reading (AMR) and 'smart' meters.

Much of the UK government guidance provided since 1990's (BRECSU 1996; BRECSU 1998; BRECSU 1998; BRECSU 1998; The Carbon Trust 2005; The Carbon Trust 2006; The Carbon Trust 2008) recommends the collection and analysis of utility billing data as a means to generate energy management information. Billing data are a valuable resource for energy management and much can be gained from their analysis. Energy bills in the UK are typically issued on a monthly or quarterly basis. The prevailing analysis methodology is described in section 2.2.

One disadvantage of using utility bills is that they can be based on estimated readings. Estimation is acceptable for billing systems as it ensures customers receive a timely bill. However, estimation is of no use to serious energy management as any analysis or interpretation of estimated readings will reveal nothing about actual performance.

Another problem is that billing data are collected at effectively arbitrary intervals. The meter reading may not be taken at the same time each day or even on the same day each month. In order to map energy consumption to degree-day figures readings must be apportioned according to average consumption per day and reassigned to calendar months.

In extreme cases the low quality of billing data can be such that they are actually worse than useless, their analysis leading to inappropriate conclusions. Notwithstanding these common problems, it is sometimes the case that utility billing data can be of very trustworthy quality and can lead to a valuable analysis. Most commonly, billing data require some apportionment but are 'good enough' to conduct a useful analysis.

Where problems with billing data are insurmountable a simple solution is to collect meter readings manually. Not only does this produce a reliable dataset for energy management information, it can also be provided to the utility company to produce more accurate bills.

Manually reading utility meters can produce monthly, weekly or even daily consumption data. However, it is very resource intensive and, when managing large numbers (hundreds) of buildings, the logistics can become unworkable.

Since the introduction of competition into the UK electricity supply market short interval (in particular half-hourly) time-series data have become available for large buildings. Indeed half-hourly data is a prerequisite for all buildings on maximum demand tariffs (where the customer pays a premium based on the maximum load experienced in any half hour of each month) in the UK.

Prior to 1990 a typical building would have 4 quarterly, 12 monthly or 52 weekly meter readings per meter per year. A half-hourly metered site generates 17,520 readings per year. In 1990 in the UK there were 4000, mostly industrial, sites metered for half-hourly data. In 1994 there were 12,000, by 1998 there were 60,000 and by 2002 this had risen to 100,000 sites of all types (Stuart, Fleming et al. 2007).

Due to recent developments in areas such as communications technology automated meter reading (AMR) systems are now cheaper and a more realistic proposition for smaller sites. This enables their use as the primary source of energy management information in multi-site organisations such as local authorities. A comprehensive guide on building energy metering is provided by (CIBSE 2009).

Article 13 of the Energy End-use Efficiency and Energy Services Directive of the European Union (EEEESD) requires member states to ensure that final customers for energy have meters that provide actual consumption levels and information on actual time of use (Warren 2003). The EEEESD also requires bills to be provided “frequently enough to enable customers to regulate their own energy consumption” and requires comparisons with historical consumption and with benchmarks based on the normalised consumption of users in the same category (EU 2006).

High-resolution energy consumption data are likely to be more common in the future. The volume of these data will only grow and an analysis methodology such as that developed in this work is likely to be necessary to process them automatically.

3.2 Data collection

The buildings under analysis in the present work all belong to or are run by Leicester City Council. The varied building stock is spread across the city of Leicester and includes swimming pools, schools, offices, libraries, museums and district heating for housing estates. Monitoring this geographically dispersed building stock requires a scalable metering solution.

Leicester City Council has made significant investments in a city-wide data collection infrastructure which was the first of its kind for a local authority in the UK. Data are collected automatically through a communications network linking over 1000 utility meters. This section introduces the system and briefly describes the data it has generated to date.

3.2.1 Communications network

The network comprises an infrastructure of metering and communications hardware. All meters are connected to a transmitter and seven receiver nodes are strategically located around Leicester city so as to provide an umbrella within which there is always a receiver in range.

As illustrated in Figure 3.1 the network relies on seven radio receivers, located strategically around Leicester city. The basic 'umbrella' provided by the overlapping coverage of these seven receiver ‘nodes’ means nearly every building within the city centre has a receiver within range and so meters can be added easily. Temperature data are recorded at the point indicated on the figure.

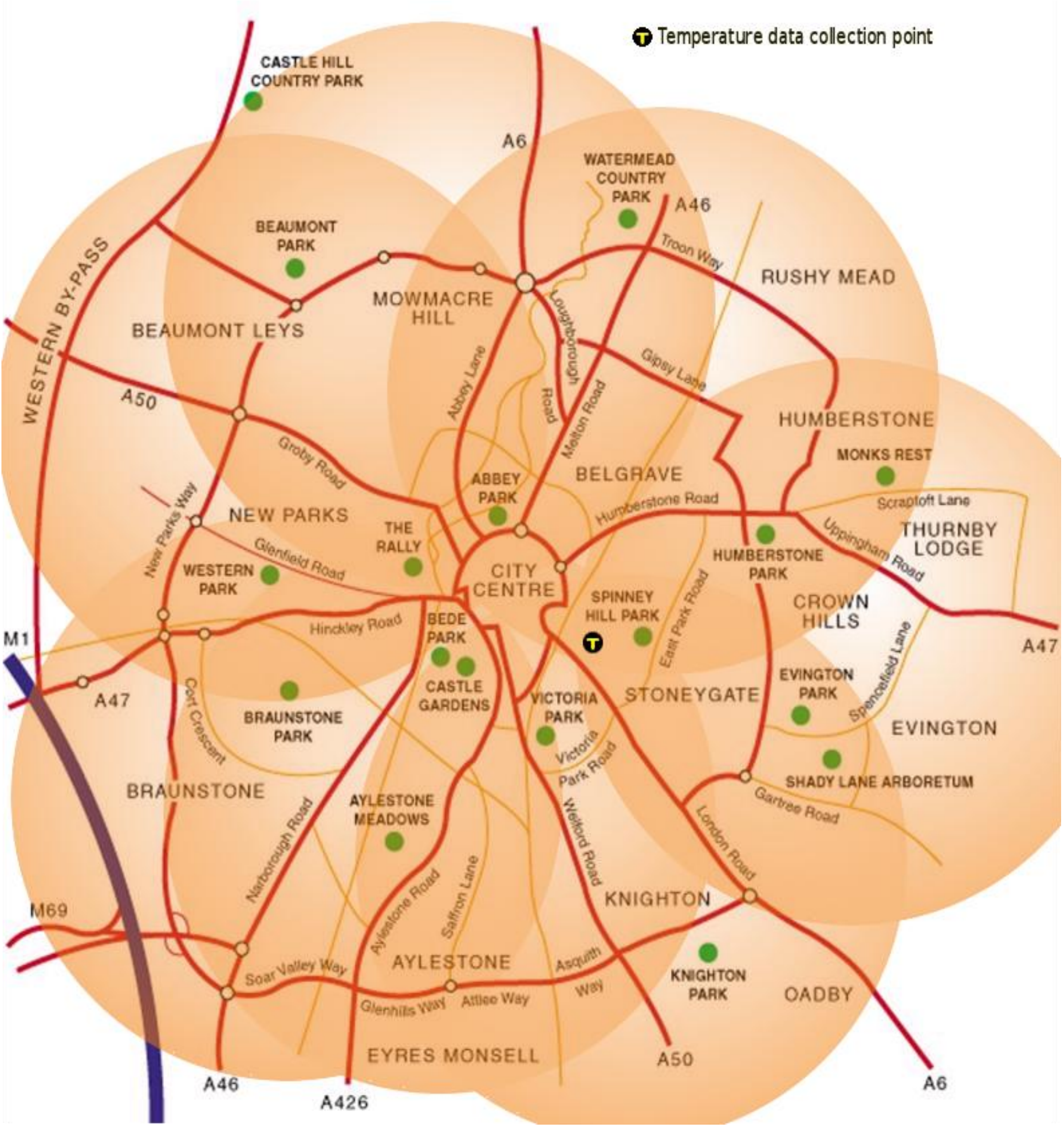


Figure 3.1: Map of Leicester city showing coverage of seven receiver nodes

Figure 3.2 shows the basic system structure. Meters generate consumption data and transmit the data to one of the receiver nodes. The central server interrogates each receiver node regularly to import data.

Each meter or collection of meters is connected to a radio transmitter. Transmitters are tuned to the frequency of their nearest receiver. The data are transmitted along with a unique identifier to indicate the data source. Receiver nodes record the data from several meters.

Specialised software on the central server regularly interrogates the seven main nodes and imports any newly generated data. This process is conducted on a daily basis thus usually making data available in the central database within a day.

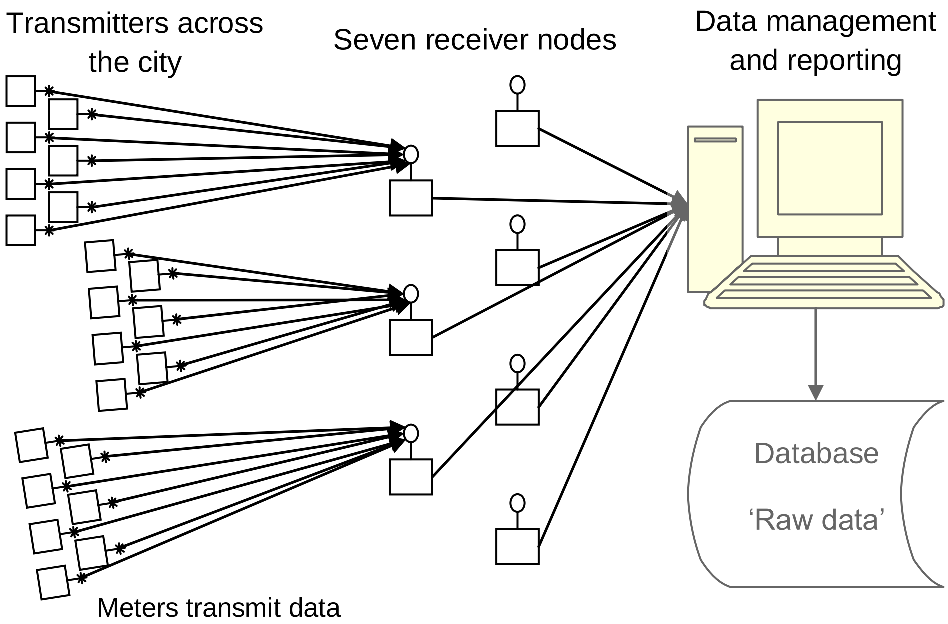


Figure 3.2: Basic network diagram

Several different kinds of meters are connected to the system including gas, electricity and water meters. In addition to energy consumption, the system records battery voltages for standalone data transmitters and is also connected to an outside air temperature sensor located centrally in Leicester.

New meters can be easily linked to the network by installing a new radio transmitter. Where several individual meters are dispersed across a large building they can be connected to short distance transmitters which relay their data to a single long-distance transmitter in the building.

3.2.2 Configuration

The resolution of the automated data collection systems can be configured. The typical arrangement is that a meter reading is collected every half hour. It is often the case that during the commissioning of a given meter, data are collected at a higher frequency for a short period to confirm that data are being collected correctly.

Many hundreds of meters are communicating their consumption data via low-powered radio every half hour across Leicester. Though data are collected every half hour they are not collected precisely on the half-hour. Each dataset is collected at a random offset determined at the time of installation. This spreads the load on the receiver nodes over time and avoids any problems with interference or receiver saturation.

Automatic meter reading hardware is connected directly to the pulse output of fiscal meters. Calibration information (i.e. the value to attribute to each pulse) is recorded in the central database against each meter code. During the commissioning process, a pair of manual meter readings is taken, several days apart and compared to automatic readings to ensure the system is correctly calibrated. This provides a degree of confidence that the meters are correctly calibrated and that the data provide an accurate reflection of consumption.

Temperature data is only collected at a single point in the city of Leicester and is not set up in a proper weather station. This may be a limitation as it may be expected that many buildings will have their own micro-climate determined by their immediate surroundings and unless the sensor is very carefully located, it could be affected by solar radiation.

Also, since the data are collected at only one point, the quality of this temperature dataset will impact all the analyses in which it is used. The temperature sensor was carefully located in the city centre so is considered representative of the conditions experienced by most of the buildings under analysis.

3.2.3 Descriptive statistics

For use in this work a snapshot of the database was taken in April 2008. The full database includes a total of nearly 80,000,000 meter reading records (i.e. rows in a database table). This large dataset includes data from 764 gas and electricity meters but also includes data from water meters and other data that will not be analysed as part of this work. Table 3.1 includes some basic statistics regarding the number of meters.

Table 3.1: Main dataset - basic statistics

Quantity	Value
Number of sites	310
Number of gas meters	329
Number of electricity meters	435
Number of meter readings	47,898,107
Average readings per meter	62,367

The full dataset under analysis includes 47,898,107 gas and electricity meter reading records. A simple calculation shows that this is an average of 62,367 readings per meter. If the data are assumed to be half hourly the assumption is that there are 17,520 readings per year, thus the number of data records indicate an average of about 3.6 years of data per meter.

The data cover a period of over seven years with the first meters being installed in early 2001 and the longest running datasets covering the entire period through to April 2008. Of the 764 meters, 710 were ‘live’ and still collecting data in April 2008. Several of the datasets ended before April 2008 due to buildings being closed or sold or otherwise losing their connection to the data collection system.

Figure 3.3 shows how the system has grown over time. These figures were compiled from noting the date of the first and last meter readings available for each meter. The columns indicate the number of additional meters connected in each month while the area shows the cumulative total number of connected meters. The system has grown from around 50 meters in 2001 to over 700 meters in 2008. It is also possible to see distinct periods of growth in 2003 – 2005 and particularly in the winter of 2005 – 2006.

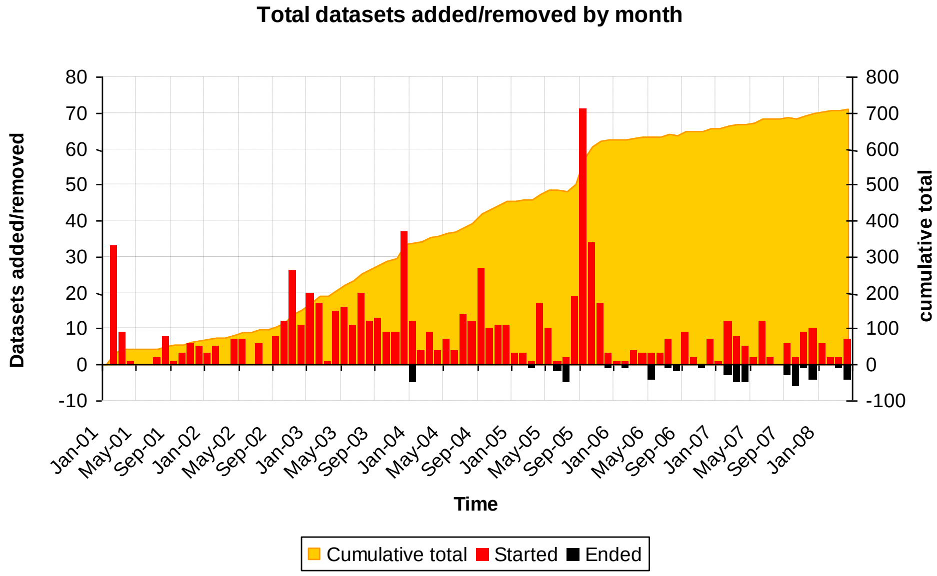


Figure 3.3: Total number of datasets (cumulative)

This gradual growth means that there are more datasets available for more recent years and only relatively few meters have more than five years of data available. This will be apparent in the analysis described in sections 5.3.1 and 6.1.4. However, it is clear that this dataset is very large and that it represents a significant information resource.

The data cannot be easily analysed in their raw form. They must be extracted from the database and imported into custom software to implement the algorithms developed as part of this research.

3.3 Research tools

In this research the raw data produced by the LCC metering network were managed and analysed using a database coupled with a programmable array-oriented data analysis language. The main software selected for these tasks were mySQL and IDL (Interactive Data Language). These tools provided the flexibility to work with large datasets and arbitrarily complex algorithms needed to develop the research presented in this work.

3.3.1 Data format

The proprietary software (Dynamat 6) used to collect and import data from the metering network described in section 3.2.1 was written using a variant of the dBase programming language. Such systems manage data in the dBase file format (*.dbf).

Several files were provided, each representing one database table. These contain the core meter reading data plus several tables of supporting data such as meter names and calibration settings.

The metering data themselves are stored in a table named `READ_FIL`. This includes many millions of records including all the data from every meter in the network. The key fields in this table are the timestamp (split into date and time fields), the cumulative meter reading (referred to as `INTEG`) and a reference number which indicates the source of the data (referred to as `KEYNAME`).

A further table, `KEYNAMES` includes a list of meters with supporting information such as meter locations, meter name and calibration settings. The `KEYNAMES` table also includes the `KEYNAME` field which uniquely identifies each record such that the tables can be linked.

The calibration settings allow for the meter pulses recorded in `READ_FIL` to be converted into common energy units. In this work the majority of figures are quoted in kWh rather than SI units as this is the most common unit used in industry.

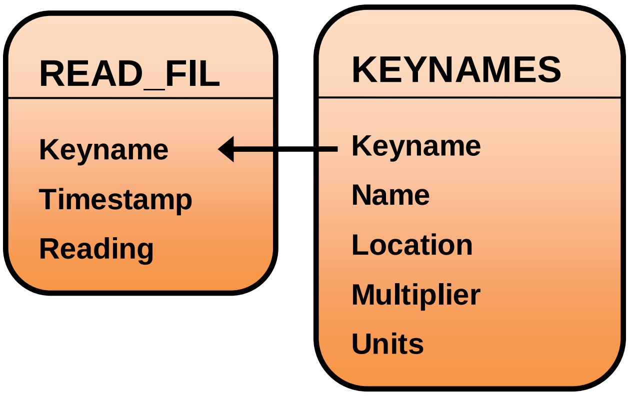


Figure 3.4: Basic database structure for metering data

Figure 3.4 illustrates the simple relationship between these main tables and summarises the most important fields.

3.3.2 mySQL Database

Relational data separated into tables allow for complex, structured data to be stored efficiently with a minimum of redundancy. The database format is also convenient because it can handle very large volumes of data and specific subsets can be extracted conveniently via SQL queries.

To manage the data effectively the .dbf database files were imported into a mySQL database. Initial tests indicated that data access was very slow (taking up to a minute to extract a single dataset from the database). Some basic optimisations improved performance considerably. The optimised system returned an average dataset in a matter of two or three seconds.

Repetitive data such as units and meter type information were normalised to remove redundancy. Database normalisation involves splitting one or more fields from one table into a new, more compact, lookup table.

For instance, rather than storing the text string related to the unit type (e.g. ‘kilowatt hours’) several thousand times, each string is given a unique integer identifier and the integer is stored against the appropriate records. Since keys are integers they take less space and are more easily compared.

Other basic processing included removing unnecessary fields, replacing the two timestamp fields with a single field and replacing the text string key with an integer. Indexes were added to the tables to allow for fast queries. This all served to reduce the size of the data without losing any important information, thus speeding up access to the data of interest.

The mySQL database represents a crucial part of the research tools used in this work. Not only was it used as the main source of ‘raw’ data and to perform basic analysis via SQL queries, it was also used to store and analyse the results of analysis.

3.3.3 Analysis tools

The mySQL database provides a convenient repository for the raw data but cannot provide the complex data manipulation required to conduct a complete analysis. A programmable tool capable of complex arbitrary data transformations was required.

The IDL programming environment was selected as the main analysis tool. Key benefits of IDL include fast array-based data manipulation and an object-oriented syntax. Any of several similar tools would have been just as appropriate (e.g. R, Matlab or Python) but the availability of expertise and support within the workplace meant that IDL was the most appropriate choice for this researcher at the time.

Linking IDL to mySQL involved developing a database toolkit for IDL whereby SQL queries were passed to the command line mySQL client and the resulting text string parsed into a native IDL array. Though this was not the most efficient approach, it sufficed to provide a means to query the database and access the metering data from within the IDL environment via a simple function call.

Some simpler analysis was also conducted in spreadsheets. Both OpenOffice.org Calc and Microsoft Excel were used to develop the basic analyses for this work.

3.4 Metering data

Each dataset contains a simple time series of meter readings. This section describes an individual dataset and the processes required to convert meter readings into energy consumption data.

In the main, the data quality is good but two significant data quality issues have been identified. Missing and irregular meter readings leading to a variable time step cause difficulties with mapping datasets to temperature data. Erroneous readings resulting in steps and spikes in the data cause bias in consumption models.

It should be noted that of the nearly 48 million readings only very few are in error and the occurrence of larger chunks of missing data is also rare. However, it was necessary to develop a sophisticated methodology for data cleaning and standardisation. This methodology and related processes are described in detail in 0.

3.4.1 Data format

Each record (row) in the meter readings table includes a timestamp and a pulse count. One pulse represents a fixed amount of consumption; the value of a pulse varies depending on the meter calibration. The timestamp is a value which represents the time (to the nearest second) at which the reading was taken. The pulse count represents the number of pulses the meter has recorded since it was last reset to zero.

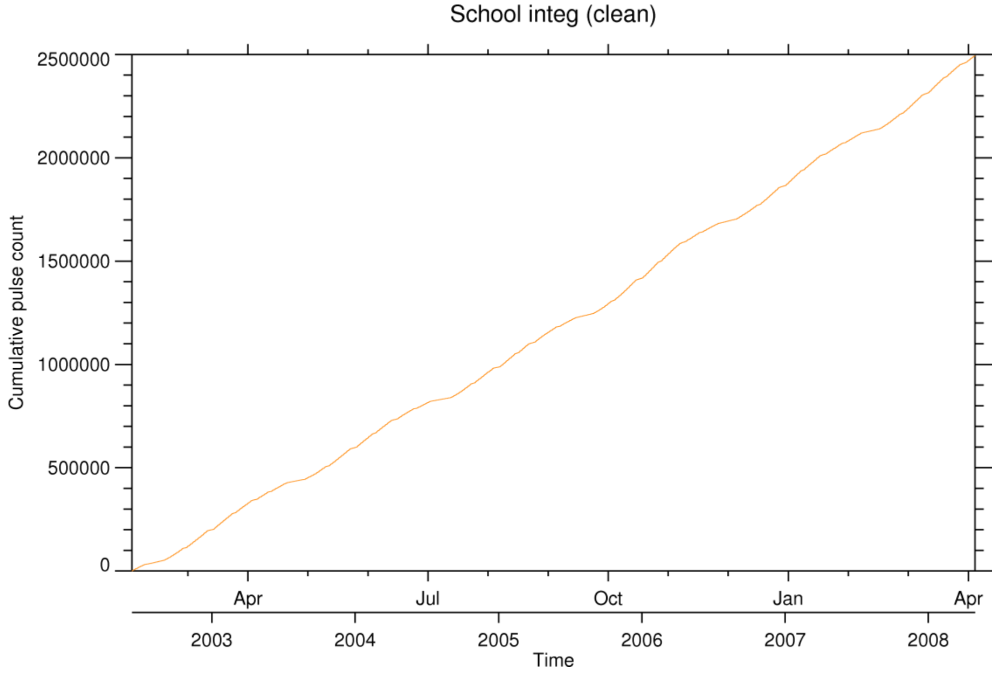


Figure 3.5: Example meter reading data

Figure 3.5 shows an example dataset (in this case, electricity consumption in a school) covering nearly six years. The dataset includes 116,159 individual meter readings. The pulse count increases from zero at a rate proportional to the rate of consumption. Some features can be identified directly from this data. The seasonal nature of consumption in this building (related to school holidays) causes the gradient to drop significantly during the summer period and other holidays.

Storing data in this cumulative format is advantageous since it provides a built-in ‘safety net’ which reduces the impact of any missing or poor quality meter readings. Figure 3.6 shows an artificial dataset created for demonstration purposes.

Meter readings (red spots) are taken at discrete points in time. For each pair of meter readings the total energy consumption can be determined. No matter how large the gap between readings, the total consumption in the period can be determined.

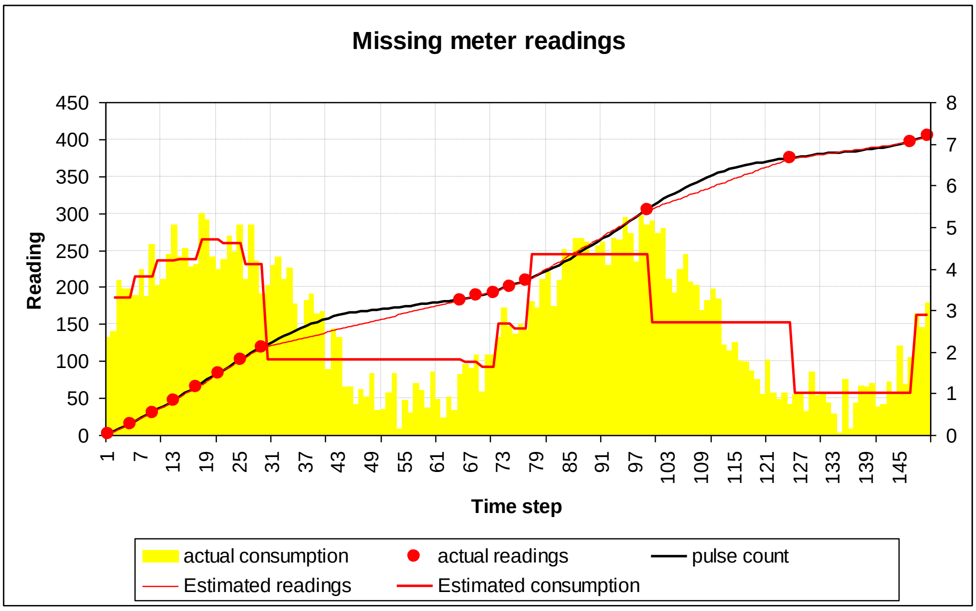


Figure 3.6: Meter readings and interpolation

The estimated consumption (based on linear interpolation, see section A.3.3 in Appendix A for details) reconstructed from the meter readings is close to reality even though the meter readings are far apart. The cumulative total is always correct at the point where a meter reading is taken.

If a meter reading error occurs then the cumulative total is recorded incorrectly for a single reading. This causes errors in the calculated consumption for the periods beginning and ending with the erroneous reading.

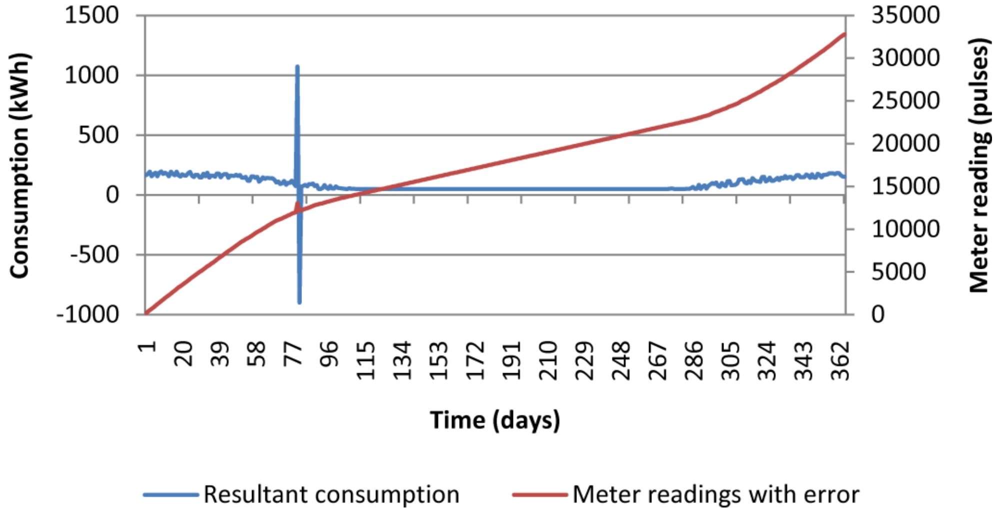


Figure 3.7: Meter reading error

However, as long as the next reading is correct, the sum of these errors will be equal to the correct consumption for the period. This means that, even if many readings are in error, as soon as a correct reading is made the running total of consumption will be correctly determined.

Figure 3.7 shows how an error in meter readings can generate consumption data with equal and opposite errors. These cancel each other out when summed. Errors like these, if sufficiently large, can be identified and removed automatically (see 0).

Critically, with a cumulative dataset such as this, missing readings do not corrupt the entire dataset. The resolution is reduced but the remaining readings are still correct. Conversely, if the absolute consumption is recorded (see Figure 3.8) for each period between readings then if data were lost or incorrectly recorded any calculation of the total consumption for a period would be affected.

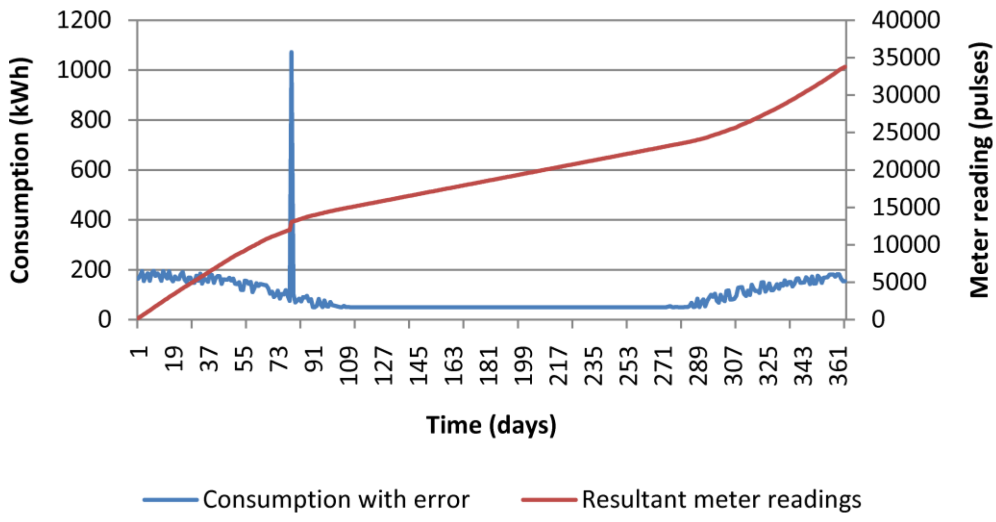


Figure 3.8: Consumption error

The utility industry relies on this convention to ensure customers are eventually billed correctly even if bills need to be estimated in the short term. It does not matter how long the period between actual meter readings is, once a reading is taken, consumption since the last reading can be calculated and an accurate bill can be issued. Equally, if a meter reading (or estimate) is inaccurate then the next reading can be used to rebalance the billing (and issue a rebate if necessary).

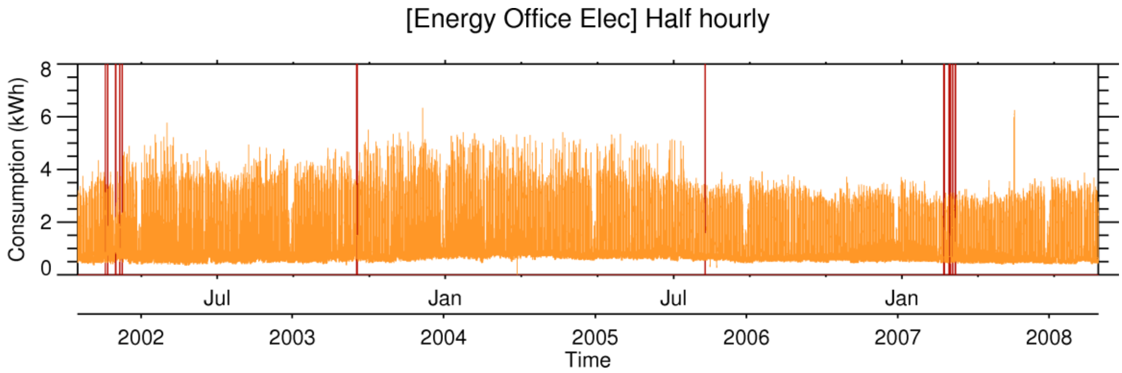
The cumulative data format provides a robust system but the cumulative nature serves to hide the detail. Figure 3.5 shows that the raw meter readings clearly contain information about energy consumption patterns but they are difficult to analyse visually. The data must be processed before they can be effectively visualised and analysed.

The main problem when dealing with this data is one of processing the data in such a way as to extract the key information. The data must be formatted and assessed for quality before they are prepared for the modelling process itself. This process is very complex (and quite boring) and so is described separately in 0.

3.5 Data preparation

The analysis is conducted at a daily resolution with raw data being interpolated to daily total consumption and average daily temperatures. Daily analysis reduces the complexity of the method greatly. The detail of precisely when an event occurred to the minute is not very important for energy management on a large scale. Figure 3.9

shows both half hourly and daily interpolated electricity consumption data from the offices of the energy team at Leicester City Council.



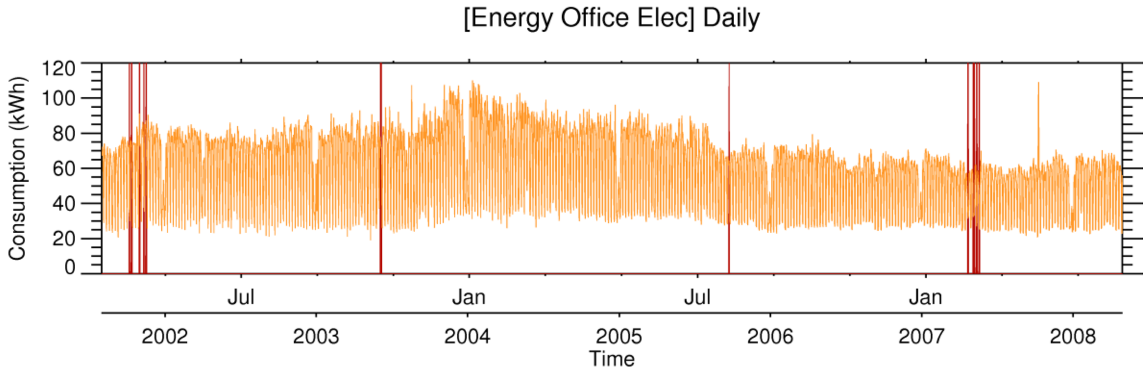


Figure 3.9: Interpolated data (half hourly and daily)

During interpolation to daily data, missing data are absorbed into the interpolated values. Where missing data are identified (that is, periods greater than 24 hours between readings) the daily values which are affected are flagged and not included in further analysis. The missing data for the energy office are highlighted in the figure. It is apparent that in this case the missing data has little effect on overall data quality. More detail on data quality and the data cleaning processes used in this work can be found in 0.

Chapter 4 Methodology

“On those who step in the same river, different and different waters flow”

“All things move and nothing remains still”

– Heraclitus the obscure (c. 535 – c. 475 BCE)

This chapter presents the detailed method developed in this work. The approach is based on event detection using OLS CUSUM (see section 2.5) and a set of energy consumption models based on the VBDD model (see section 2.2). Consumption models and event-detection methods are described and demonstrated using real data derived from raw metering data introduced in Chapter 3.

In this work it is assumed that building energy consumption patterns are consistent in the short term and can be represented by simple mathematical models. It is also assumed that changes occur to these consistent patterns. Indeed, there is a strong assumption in this work that changes to consumption pattern occur abruptly, over a short period of time. As such this work aims to represent changing patterns of consumption as a series of consumption models applied to discrete periods separated by events.

This chapter is organised into two main sections describing the modelling and event detection methodologies. In order to capture consumption patterns in many different datasets, a collection of alternative consumption models are used. These models and an automatic method for selecting the most appropriate model are described in section 4.1. Multiple events in a single dataset are detected by applying OLS CUSUM as a binary recursion. The approach is described in section 4.2.

4.1 Consumption models

In this work consumption models are intended for the analysis of daily resolution consumption data such as that shown in Figure 3.9. As discussed in section 2.2, consumption models can be used with data at lower resolutions and could apply to higher resolution data also. In this work, for event detection and data analysis, daily resolution is chosen for speed of analysis and as a compromise between simplicity and precision.

Energy consumption in buildings can follow a variety of occupancy patterns and may or may not be influenced by weather. A ‘one model fits all’ approach could be taken but would lead to a complicated model being used to cover all circumstances. For many datasets, a simple alternative would not only provide just as good a fit to the data but would also describe consumption in a more sensible way. The consumption model employed in this work actually comprises eight basic alternative models representing two core models described in section 4.1.1 and four common consumption patterns described in section 4.1.2.

In the interests of parsimony it is desirable to use the simplest model possible to capture the variation in a given dataset. Methods for determining model goodness of fit are described in section 4.1.3. In this work, the addition of complexity (i.e. increasing the number of model parameters) can only be justified by a commensurate increase in model predictive power. As such the modelling process consists of fitting several models to a given dataset and choosing the most parsimonious alternative. An automatic model selection method is described in section 4.1.4.

4.1.1 Core models

There are two core models used in this work. They have been chosen for their simplicity and their applicability to the dataset under analysis. The type 1 model is a simple constant model which is appropriate for datasets where outside air temperature has no influence over consumption. The type 2 model is the standard VBDD model introduced in section 2.2.1 which is appropriate for datasets which vary with outside air temperature.

The constant (type 1) model

In many cases there is no connection between outside air temperature and consumption. The most basic model is referred to in this work as the ‘type 1’ model. According to this model, every day is effectively the same in terms of energy consumption. The type 1 model is suitable for many datasets, in particular non-heating consumption in buildings which are occupied every day of the year.

The model takes the simple form

$$\hat{E}_{1,t} = f_{nh} \tag{4.1}$$

where $\hat{E}_{1a,t}$ is the predicted consumption according to the type 1 model at time t and f_{nh} is the fixed consumption unrelated to heating. The value of f_{nh} is estimated very simply as the mean of all the measured data in the period of interest.

$$\hat{f}_{nh} = \frac{1}{n}\sum_{i=1}^{n} E_i \tag{4.2}$$

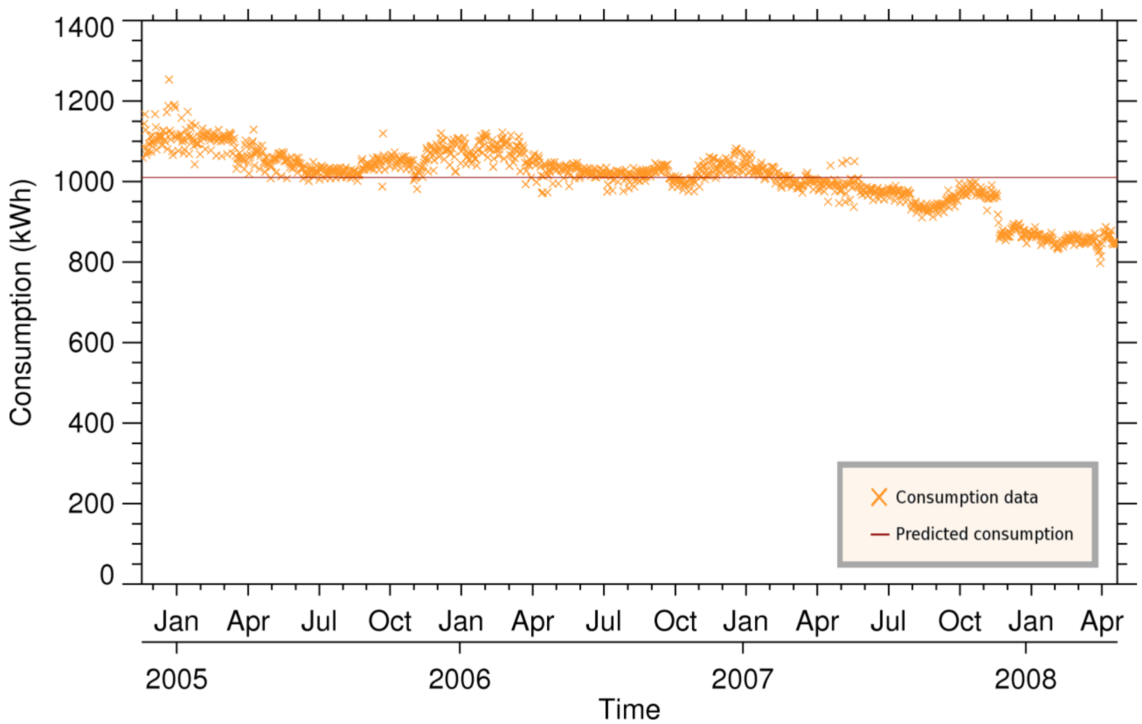


Figure 4.1: Example of the type 1 model prediction

Figure 4.1 shows the prediction generated by the type 1 model fitted to over three years of daily electricity data from a car park. The data are generally constant in that there is no obvious seasonal component. However, there is a clear downwards trend which the model does not reflect. A key element of this work is the assumption that this kind of movement in consumption patterns is inevitable over long periods. After three years the average rate of consumption in this dataset has fallen by at least 300 kWh per day.

The latest consistent pattern, resulting from a clear drop in consumption in late 2007, is far below the global average but seems to conform broadly to the constant model in the short term. In this work the assumption is that events such as this can be built into the modelling process and reflected in the results of analysis.

The VBDD (type 2) model

When a dataset includes consumption for space-heating the VBDD model (referred to henceforth as the ‘type 2’ model) is appropriate. Consumption is assumed to be equal to some constant value plus (during sufficiently cold days) a constant amount of energy for every degree below a specific threshold temperature. This relationship was derived in section 2.2.

The predicted consumption at time t, $\hat{E}_{2,t}$ is defined as follows.

$$\hat{E}_{2,t} = f_{nh} + \beta\left(\tau - T_{out,t}\right)^{+} \qquad \textbf{4.3}$$

Where f_{nh} is the constant, non-temperature related consumption, β is the heating coefficient, τ is the change-point temperature and $T_{out,t}$ is the daily average outside air temperature at time t. The superscript positive sign indicates that the bracketed expression should be set to zero if it is negative.

These parameters are estimated for a given dataset using OLS linear regression (see equation 2.3). Parameter τ is set using a grid search (see section 2.2) to find the value which minimises the RMSE of the model. Following (Kissock, Haberl et al. 2003) the range of temperatures in the input data is split into twenty equally spaced bins. The model is fitted setting τ to the temperature value at the boundary of each bin in turn. The range covering the pair of bins either side of the best fitting value is then split once again into 20 equally spaced bins and the process is repeated within this smaller range.

The type 2 model is suitable for datasets where consumption varies with outside air temperature. In particular for datasets measuring thermostatically controlled space-heating. The example in Figure 4.2 shows a good fit to nearly six years of gas data from a residential care home. The type 2 model prediction is plotted with the raw data against both time (upper panel) and temperature (lower panel).

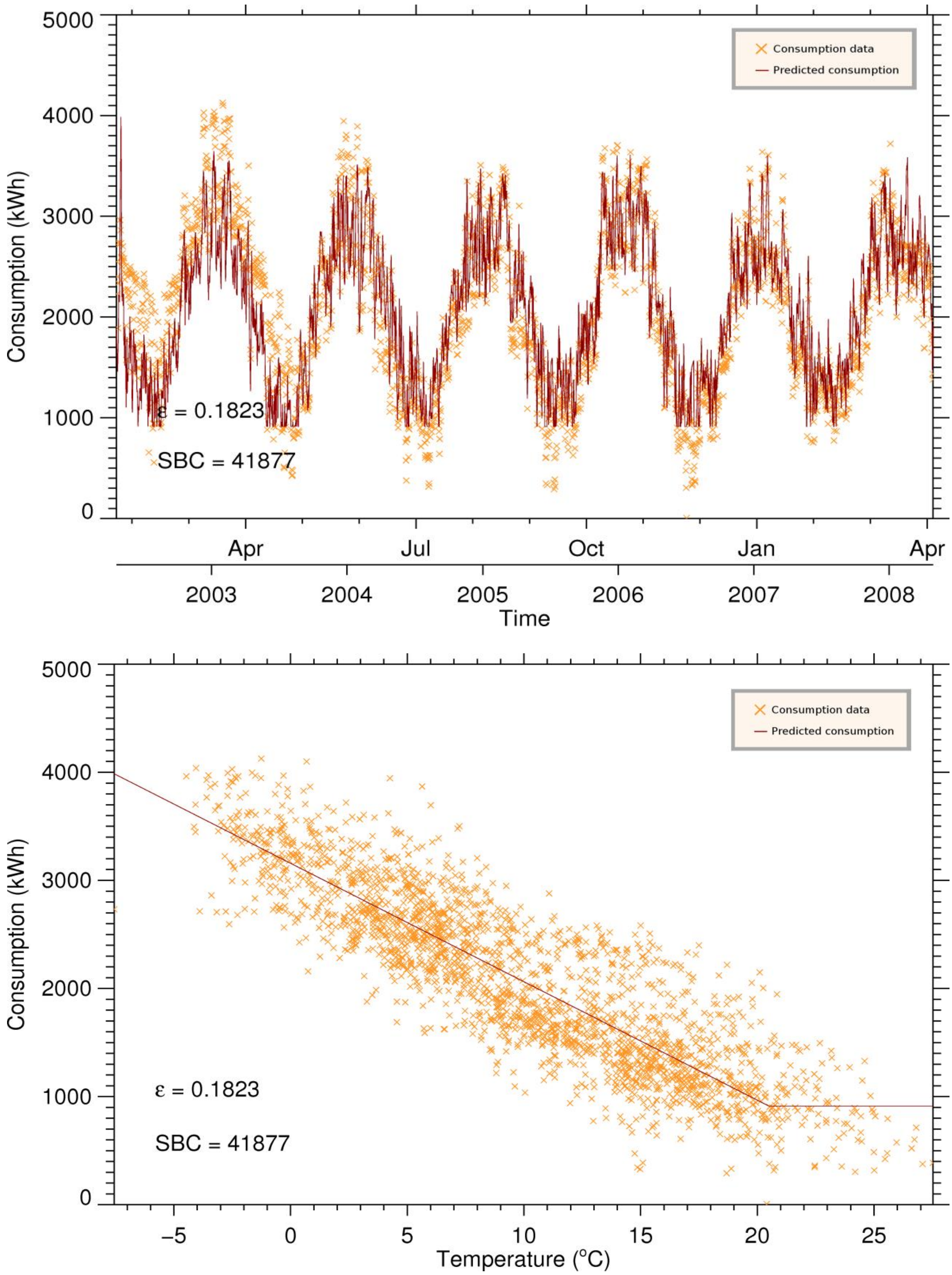


Figure 4.2: Example of the type 2 model prediction

The model provides a close approximation to the data in general. However, in 2002 and 2003 the data seem to fall above the prediction whilst beyond 2004 the pattern has dropped below the prediction. The scatter plot shows that over this period there is a wide range of consumption levels (a difference of over 1500 kWh per day between lowest and highest) for any given outside air temperature. This indicates that there may be a systematic change in the data.

4.1.2 Day of week variation

The type 1 and type 2 models cannot be used directly for all datasets as they do not take into account variations in occupancy on a daily basis. In many buildings the pattern of occupancy, and therefore the pattern of consumption, is different depending on the day of the week, a common example of this is an office building that is closed on weekends. For these buildings the core models produce very poor predictions of consumption.

This section defines variations on model types 1 and 2 which introduce a weekly element. Whilst the core models remain the same, the model fitting process is adjusted to generate weekly variants. Data are split into several subsets according to the day of the week and the given consumption model is fitted to each subset in turn. A set of model parameters is recorded for each subset.

For example, in the 'weekday/weekend' variant the data are split into weekdays and weekends. The result is two sets of model parameters; one can be used to generate predictions for weekdays; and the other for weekends. Thus, the model prediction can be reconstructed very simply.

Four such model variants are defined in this work. These include the 'base' model where all days are treated the same; the 'weekday/weekend' model already described; the 'weekday/Sat/Sun' model where Saturdays and Sundays are distinct from each other and from weekdays; and the 'full weekly' model where each day of the week is treated independently.

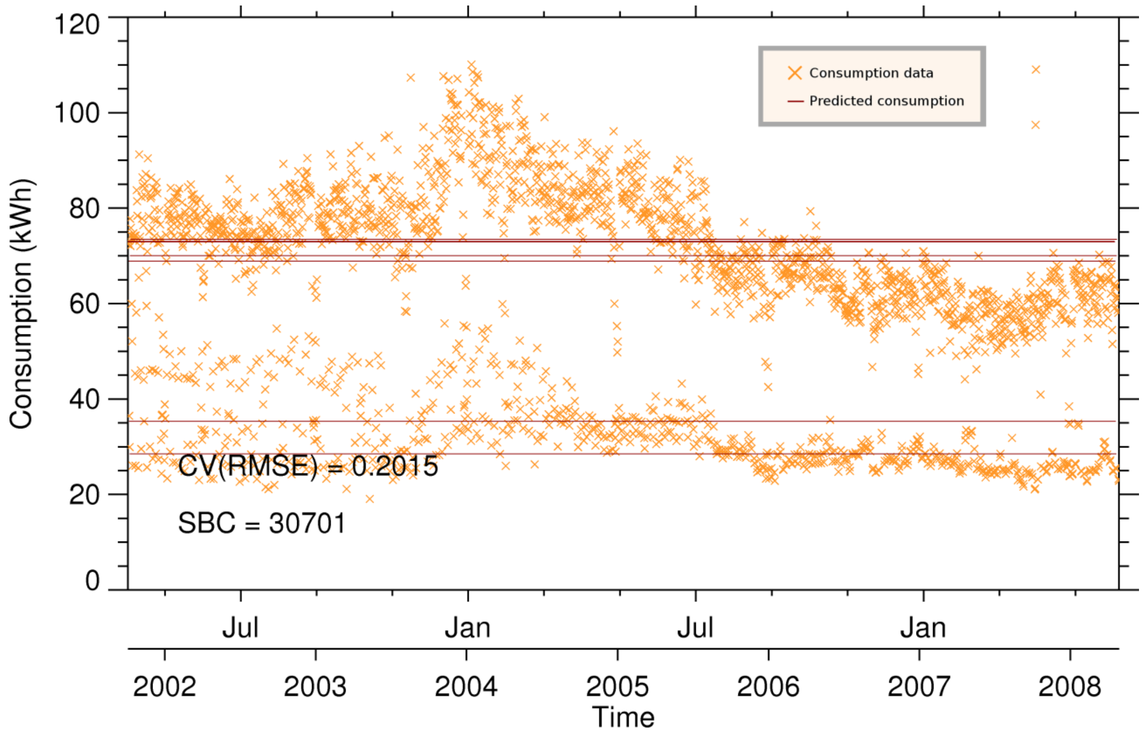


Figure 4.3: Example of the full weekly variant of the type 1 model

The example in Figure 4.3 shows over six years of electricity data from an office building. The full weekly variant of the type 1 model prediction is plotted as seven distinct lines, one for each day of the week. Though consumption is clearly not constant over the years, there is no seasonal pattern so the type 1 model is most appropriate in this case. The difference between weekdays and weekends is significant. Consumption on weekends is less than half that during the working week.

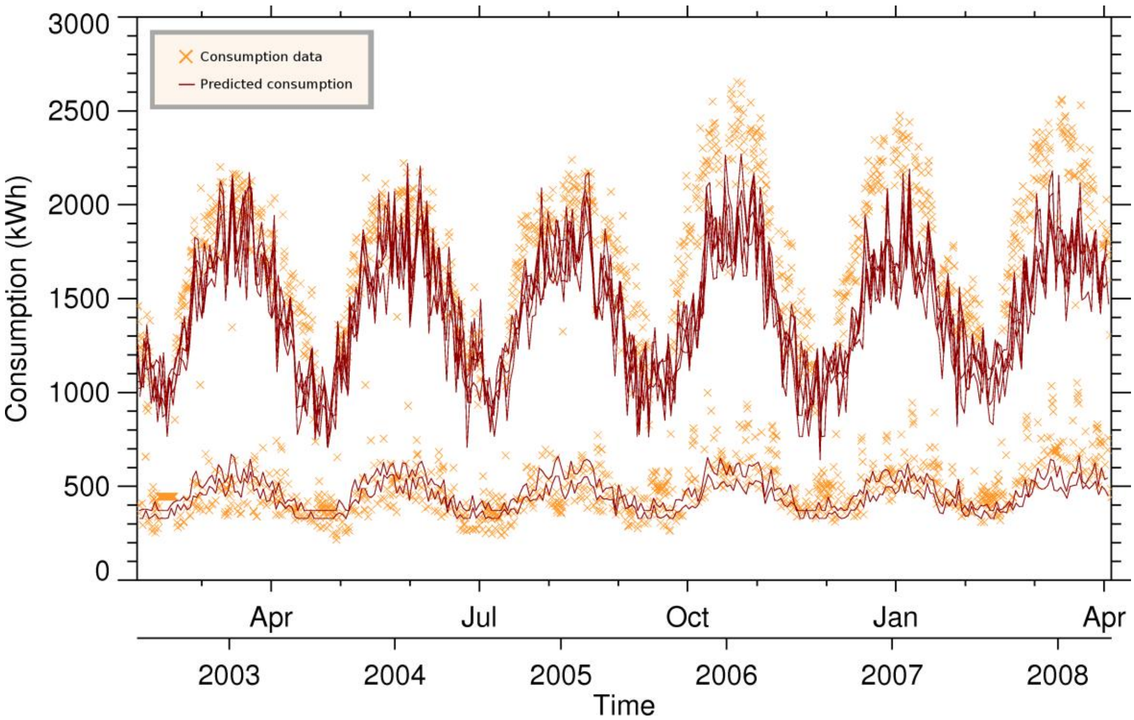


Figure 4.4: Example of the type 2b model prediction

The example in Figure 4.4 shows over five years of electricity data from an office building. In this case the full weekly variant of the type 2 model is plotted as seven distinct lines. There is a clear seasonal pattern on both weekends and weekdays but both the fixed consumption and the heating coefficient are different. It seems in this case that the weekday/weekend model may be more appropriate since Saturdays and Sundays are almost identical and the weekdays all overlap. Once again, it is apparent that there has been a major change in consumption pattern around the summer of 2005.

4.1.3 Model goodness of fit

Fitting models to data as described above will always produce a result. That is, model parameters can always be estimated even if they are entirely inappropriate for the dataset in question. This makes model fitting a dangerous business because it is possible to blindly treat model parameters as valid. To avoid making erroneous interpretations it is necessary to inspect the result to ensure it provides a suitable fit to the data.

Models are not expected to fit the measured consumption data perfectly. In many cases it is useful to have a measure of how well they fit. This is important when trying to understand if the fitted model provides useful estimates of the real consumption pattern.

In an automated analysis it is critical to have a system for isolating poorly fitting models so as to avoid erroneous interpretation. It can also be useful to compare the fit of different models to the same data.

Several metrics exist for quantifying 'goodness of fit'. They are commonly based on the model error or residuals as described by equation 2.19. Most useful for this work are the root mean squared error (RMSE) and the coefficient of variation of the RMSE. These are both described below.

Root Mean Squared Error

The root mean squared error (RMSE) is a measure very similar to the standard deviation. Indeed, for the constant model the RMSE is equal to the standard deviation. The RMSE provides a standard metric quantifying the scatter of measured consumption data around the model.

$$RMSE = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \tag{4.4}$$

Where y_i is the consumption at the i^{th} period, $\hat{y}_i$ is the model prediction for the i^{th} period and n is the number of data points. The RMSE can be calculated for any model which can be used to estimate consumption. It has the same units as the original data (in this work this is always kWh).

The benefit of using RMSE is that it can be calculated for all of the models used in this work. It is also dependent on the absolute scale of consumption which means that larger buildings with higher consumption will generally produce a larger RMSE. This makes it less useful as a general measure of goodness of fit.

Coefficient of variation of the RMSE

The coefficient of variation or CV is defined as the standard deviation divided by the mean. Alternatively, the coefficient of variation of the RMSE is defined as the RMSE divided by the mean. In the case of energy consumption models this provides an indication of the size of the error as a proportion of average daily energy consumption. Henceforth, the coefficient of variation of the RMSE will be referred to as ε to represent relative error.

$$\varepsilon = \frac{RMSE}{\mu} \qquad \textbf{4.5}$$

In this work, ε is used to indicate relative goodness of fit. That is, to determine a relative measure to indicate the extent to which a given model describes the data. High values of ε indicate high variation. This variation includes both stochastic scatter around the model and the effect of systematic behaviour not captured by the model.

4.1.4 Model selection

One aim of this modelling process is to minimise the complexity of the model whilst also maximising the goodness of fit. Fitting a model which is more complex than the data warrant provides confusing results and may lead to erroneous conclusions. Providing too many degrees of freedom in a model can lead to unhelpful results. If a model parameter is estimated to be non-zero but relatively tiny then it can serve to confuse the picture of consumption patterns unnecessarily. In some cases superfluous model parameters can lead to extreme and entirely inappropriate estimations. An example of this can be seen in Figure 4.6 below.

The analysis conducted in the present work includes a method to decide which model is most appropriate for a given set of input data. If the correct model is selected then the type of model can tell the analyst a lot about the dataset. Figure 4.5 shows an example. Four model variants are applied to a single dataset. The data clearly show a strong relationship with outside air temperature so both of the type 1 models are poor matches and have high ε values compared to the type 2 models.

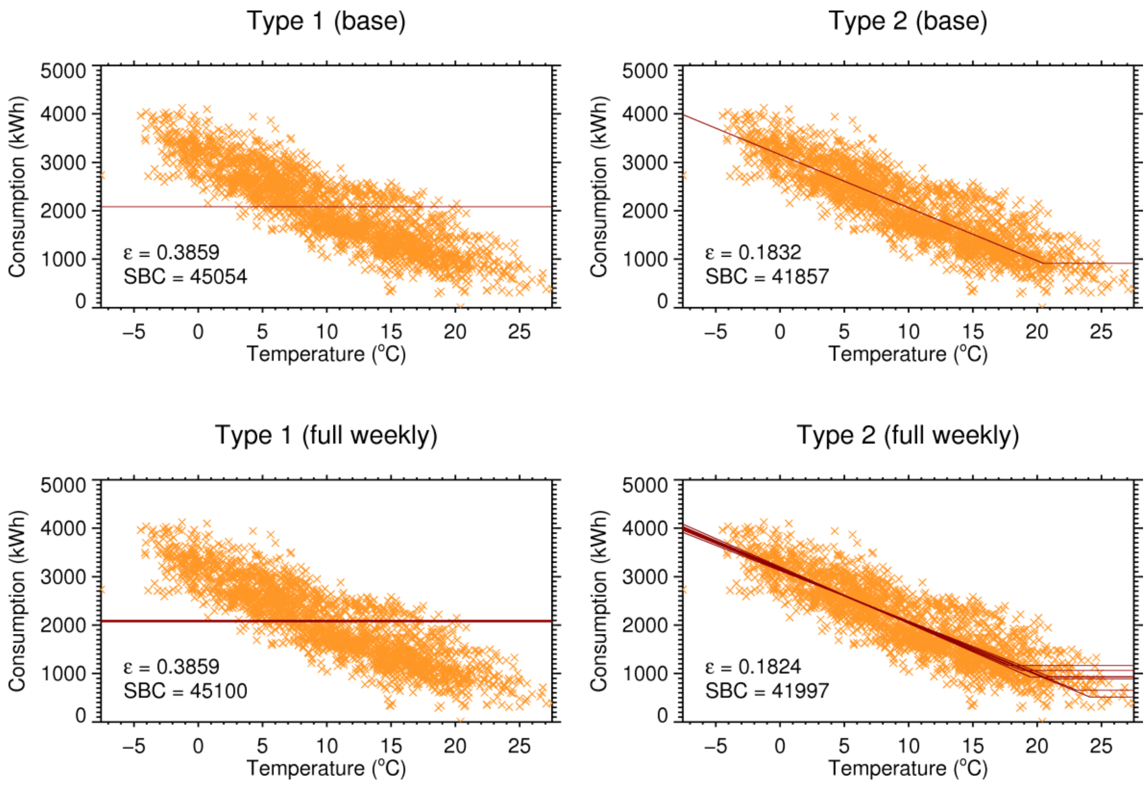


Figure 4.5: The four model variants fitted to the same dataset

Since there is little difference between consumption on different days of the week the 'base' model should be preferred. The 'full weekly' alternative is more complex than the data warrant and provides extraneous information about the tiny differences between each day of the week.

There is a problem with using the ε to assess a model in this way. Adding complexity to a model will always reduce the scatter so, ε will always be minimised with the most complex alternative. The 'full weekly' variant shows a smaller ε than the preferred 'base' variant, since it produced a 'better' fit to the data. Thus ε cannot be used to select between similar models, for choosing the most appropriate model, a different metric is required.

Schwartz Bayesian Criterion

Figure 4.5 also shows a value for the Schwartz Bayesian Criterion (SBC) for each model (Schwarz 1978). When calculated for several models fitted to the same dataset, the model with the smallest SBC is preferred. This provides a convenient way to identify the most parsimonious alternative. The SBC is larger for the type 1 models because the error is large. However, unlike with ε, the SBC is smaller for the ‘base’ variant of the type 2 model than for the ‘full weekly’ variant. This is because SBC imposes a penalty for the extra model parameters in the more complex model and this outweighs the small reduction in error it provides.

Model selection criteria such as the SBC are commonly used to choose between different models (Enders 2004). The SBC applies a penalty for each additional model parameter which must be overcome by a reduction in model error in order for the more complex model to be preferred.

The SBC is calculated according to the following formula

$$SBC = N \cdot \ln(\mathrm{SSE}) + \mathrm{p} \cdot \ln(\mathrm{N}) \qquad \textbf{4.6}$$

Where N is the number of observations, p is the number of model parameters and SSE is the sum of the squared errors (i.e. RMSE^2)

Taking the example of one year of data (N = 365) being fitted to various models it is possible to calculate the reduction in SSE necessary in each case to overcome the penalty imposed by the SBC. For a type 1 ‘full weekly’ variant (seven parameters) to be selected over a type 1 ‘base’ variant (one parameter) it must lead to a reduction in the SSE of at least 9%. For the type 2 model the equivalent reduction is 25% to select the 21 parameter variant over the three parameter variant.

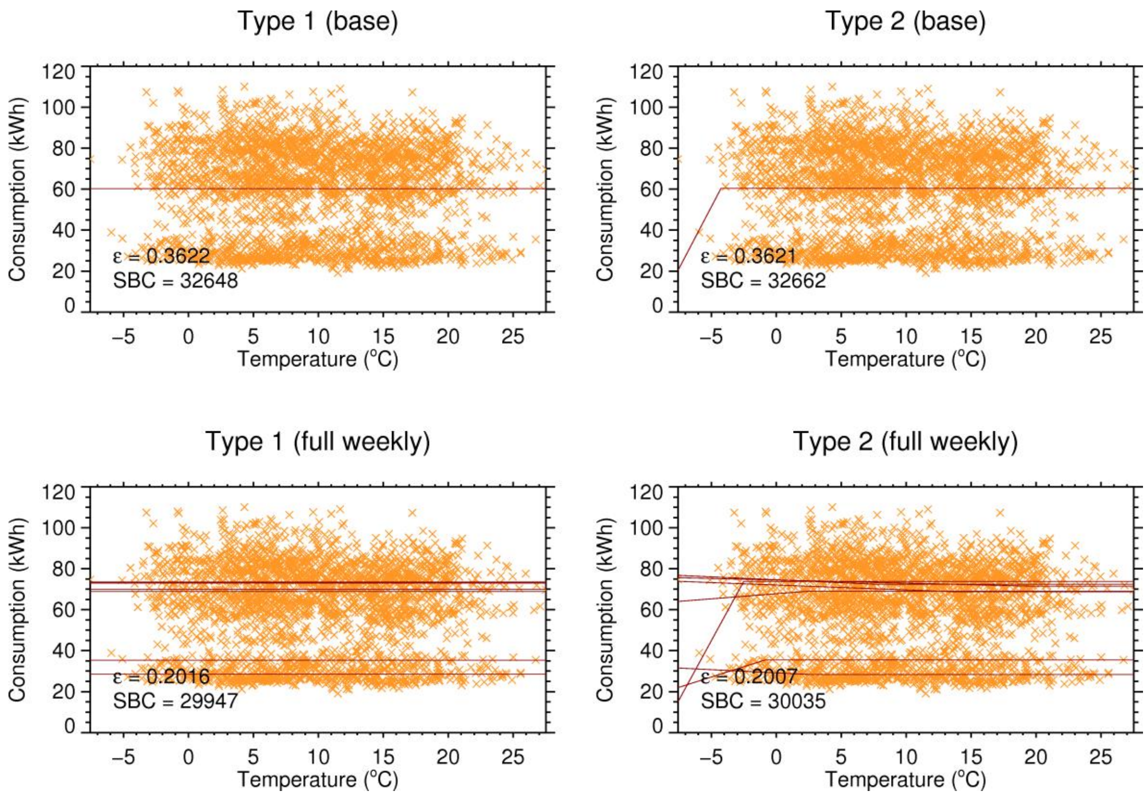


Figure 4.6: The four model variants fitted to another dataset

Figure 4.6 shows four model variants fitted to electricity consumption in an office (the same dataset as in Figure 4.3). There is no relationship between consumption and temperature but the weekly occupancy pattern has a major effect. Again, ε reduces with model complexity but the SBC is minimised for the 'full weekly' variant of the type 1 model.

Note that the models with extra, unnecessary parameters produce unpredictable results. The type 2 models estimate a change point and heating coefficient even though these parameters are meaningless for this dataset. This can be misleading and it is certainly preferable to avoid this model over-specification.

4.1.5 Model fitting algorithm

The basic model fitting process was implemented in IDL as a two-tier approach. The core models described in section 4.1.1 were implemented to be applied to any given consumption data. A third, SBC-based 'smart' model was also implemented. The ‘smart’ model simply fits both core models and chooses between them based on the SBC.

These core models are then extended into four main variants by splitting the input data into subsets according to the occupancy patterns described in section 4.1.2 and fitting the given core model to each subset in turn. From this approach the core model type and the occupancy variant can either be fixed manually or set using the ‘smart’ model selection. This produces eight standard models, comprising four variants each of the two primary core models. There are also a further four variants of the ‘smart’ core model. These ‘smart’ models are more varied however because each separate day of the week can take on either of the type 1 or type 2 models.

This leads to a very large number of potential variants. A ‘weekday/weekend’ variant could fit any of four (2^2) possible combinations of type 1 and type 2 models to the two subsets of data. For the ‘weekday/say/sun’ variant there are eight (2^3) possible combinations. If the ‘full weekly’ variant is selected then the number of possible combinations is 128 (2^7).

A final addition is the ‘smart’ occupancy variant where SBC is used to choose between the four options described in section 4.1.2. When applied with the ‘smart’ core model the SBC is used twice. In each of the four variants of the ‘smart’ model the SBC is used internally to determine (up to seven times) which of the core models to fit to each subset of the data. Then SBC is used again to select the model variant.

For example, this might result in the type 1 model being applied to weekend data and the type 2 model being applied during the week. This process is controlled entirely by selecting the model (or sub-model) which produces the minimum SBC. Using this method, any dataset can be modelled with any of the 128 combinations possible with the ‘smart’ model. This modelling process is employed extensively in the event detection algorithm described in the next section.

4.2 Event detection

Event detection is applied using OLS CUSUM as described in section 2.5.2 using the consumption models described above. OLS CUSUM allows for a given dataset and model to be tested for the presence of at least one event. The null hypothesis of no events is rejected at a given significance if the OLS CUSUM crosses the appropriate boundary.

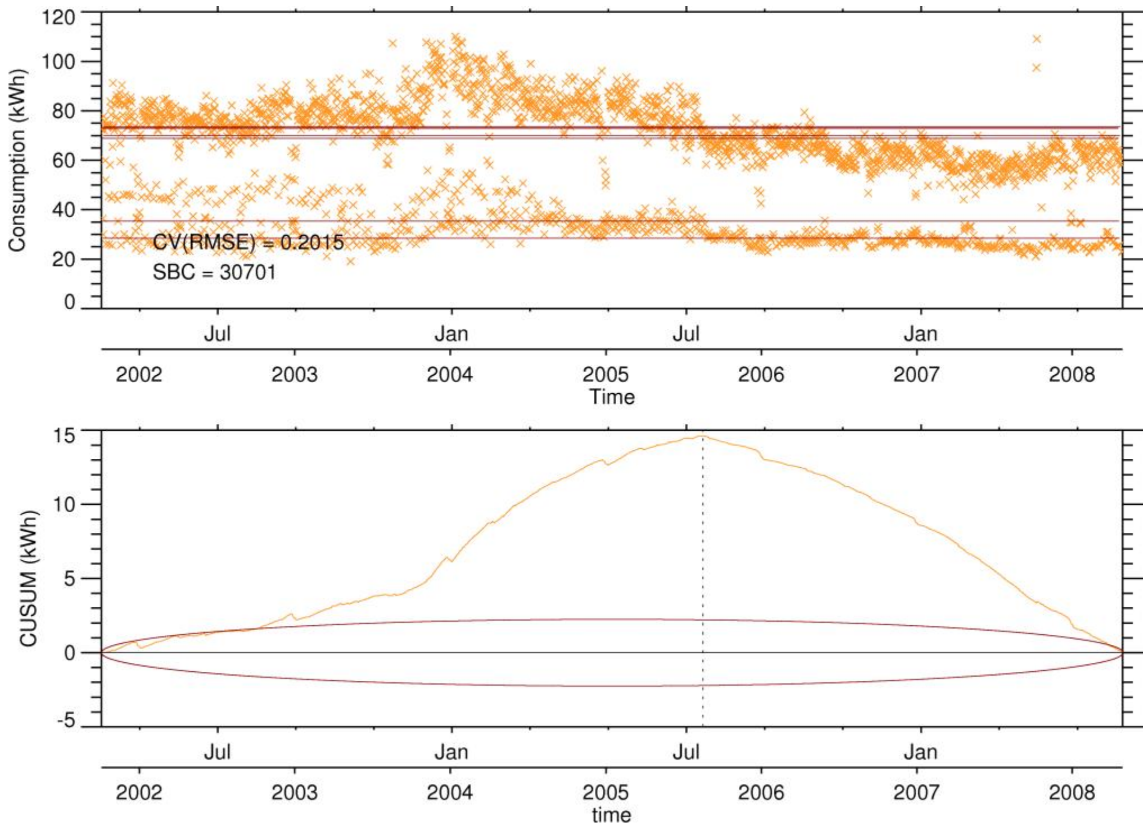


Figure 4.7: Example of OLS CUSUM with 0.1% boundary

Figure 4.7 shows the OLS CUSUM and 0.1% significance boundary for the data and model shown in Figure 4.3. The data and model are plotted again for comparison. There is a fairly obvious change in consumption levels over six years which results in a large diversion of the OLS CUSUM. The null hypothesis is clearly rejected; there is good evidence for an event in this data.

4.2.1 Event dating

Dividing the OLS CUSUM by the standard deviation of the Brownian bridge (see equation 2.23) provides a 'normalised' measure proportional to the significance of the event. This can be compared directly to the critical values given in Table 2.3.

$$\lambda_t = CUSUM_t / \sigma_{BB,t} \qquad \textbf{4.7}$$

The candidate event is identified in the figure as a vertical dotted line. It is the point coinciding with the maximum absolute value of λ. It is also the point which exceeds the boundary by the greatest proportion. If this value is above the critical value then the null hypothesis is rejected.

4.2.2 Multiple events

To handle multiple events in the same dataset this method is implemented as a binary recursion. If the null hypothesis is rejected then the data are split into two segments on the candidate event. Each segment is then remodelled and tested independently.

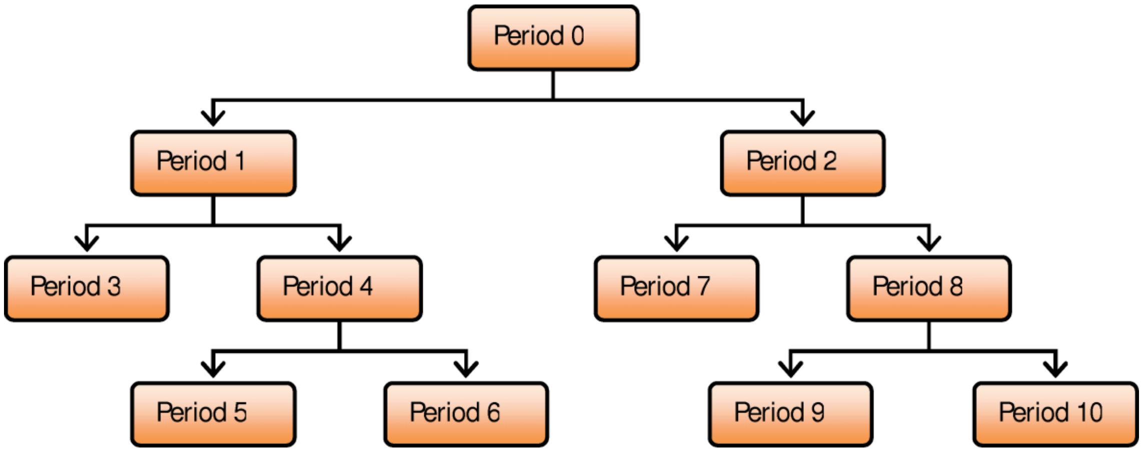


Figure 4.8: Binary recursion splits data into periods

This process continues, as illustrated in Figure 4.8, until the null hypothesis holds for each segment. Beginning with the whole dataset (period 0) the data are split into two (periods 1 and 2). Each new period is then split again until the null hypothesis holds. Period 1 is split into periods 3 and 4. Period 3 is not split further. Period 4 is split into periods 5 and 6 and so on. The final dataset is composed of those periods which were not split any further (periods 3, 5, 6, 7, 9, and 10).

Thus, the entire dataset is split into contiguous periods separated by events. The order of event detection is not the order of significance. In Figure 4.8, period 4 is split into periods 5 and 6 before period 2 is even tested. The level of significance determines when the event detection stops. If a lower significance threshold is provided then more events will be detected.

Model selection happens at each stage in the event detection process. As events are detected and the data split into segments each new segment is modelled independently. This means it is not only possible for a detected event to result in changes to the value of model parameters but also to the model type. As each event is identified its location and significance are recorded.

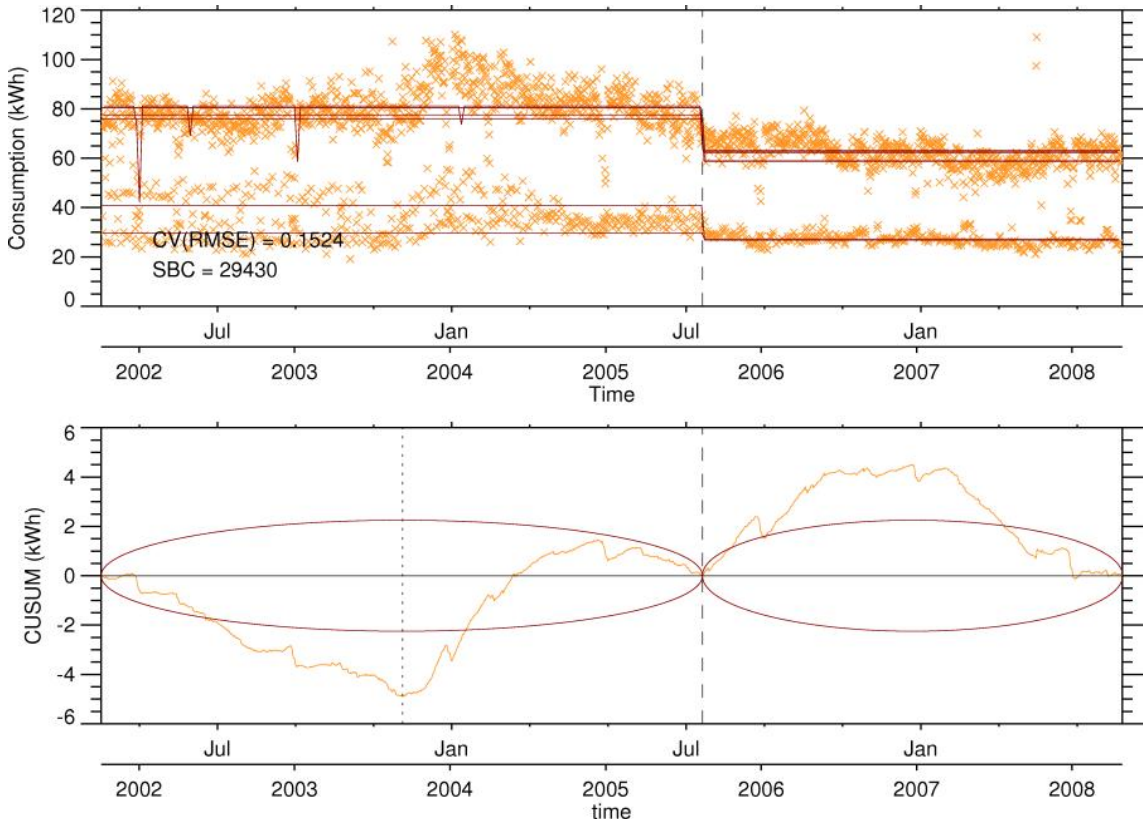


Figure 4.9: Example dataset with model and CUSUM with boundaries (one event)

Figure 4.9 shows two charts similar to Figure 4.7 but depicting the situation after the first event is detected. In the upper panel the example dataset is plotted over time with the model prediction. In the lower panel the resultant CUSUM is plotted with 0.1% significance boundaries.

The predicted consumption is the result of the two segments being modelled independently. That is, the model is fitted to the data before the event to generate model parameters for before the event and the model is fitted to data after the event to generate model parameters for after the event. The difference in these model parameters quantifies the difference between the two segments. These differences, whether increases or decreases in energy consumption, can be attributed to the event.

In both cases the model is split into days of the week. Before the event Tuesdays are fitted to the type 2 model and all other days are fitted to a type 1 model. This is due to a coincidence of cold weather and holidays. The type 1 model was rejected for Tuesdays because Christmas day caused a very large difference from expectation which was addressed in the type 2 when it was fitted with a very low change point (-2.6ºC). This type of problem could be addressed by eliminating special days such as bank holidays or including them in the modelling process.

After the event, all days of the week are fitted to a type 1 model. In general, the segment before the event shows higher consumption than that after the event. Consumption on Sundays has always been low and remains so but all other days see a drop with Saturdays falling in line with Sundays.

With this new model in place the event detection continues. The lower panel in Figure 4.9 shows the OLS CUSUM and 0.1% significance boundaries after the first event is detected. The chart shows two independent OLS CUSUM charts separated by the event. In this case the OLS CUSUM exceeds the boundaries in both cases so the null hypothesis is rejected and they will both be split further. At the event the OLS CUSUM and boundary are zero.

The process continues as in Figure 4.10, each segment is tested in turn according to the binary recursion algorithm. If the null hypothesis is rejected then the data are split into two further segments and these are tested in turn. When the null hypothesis holds for each period then no more events are detected and the final version of the model is reached.

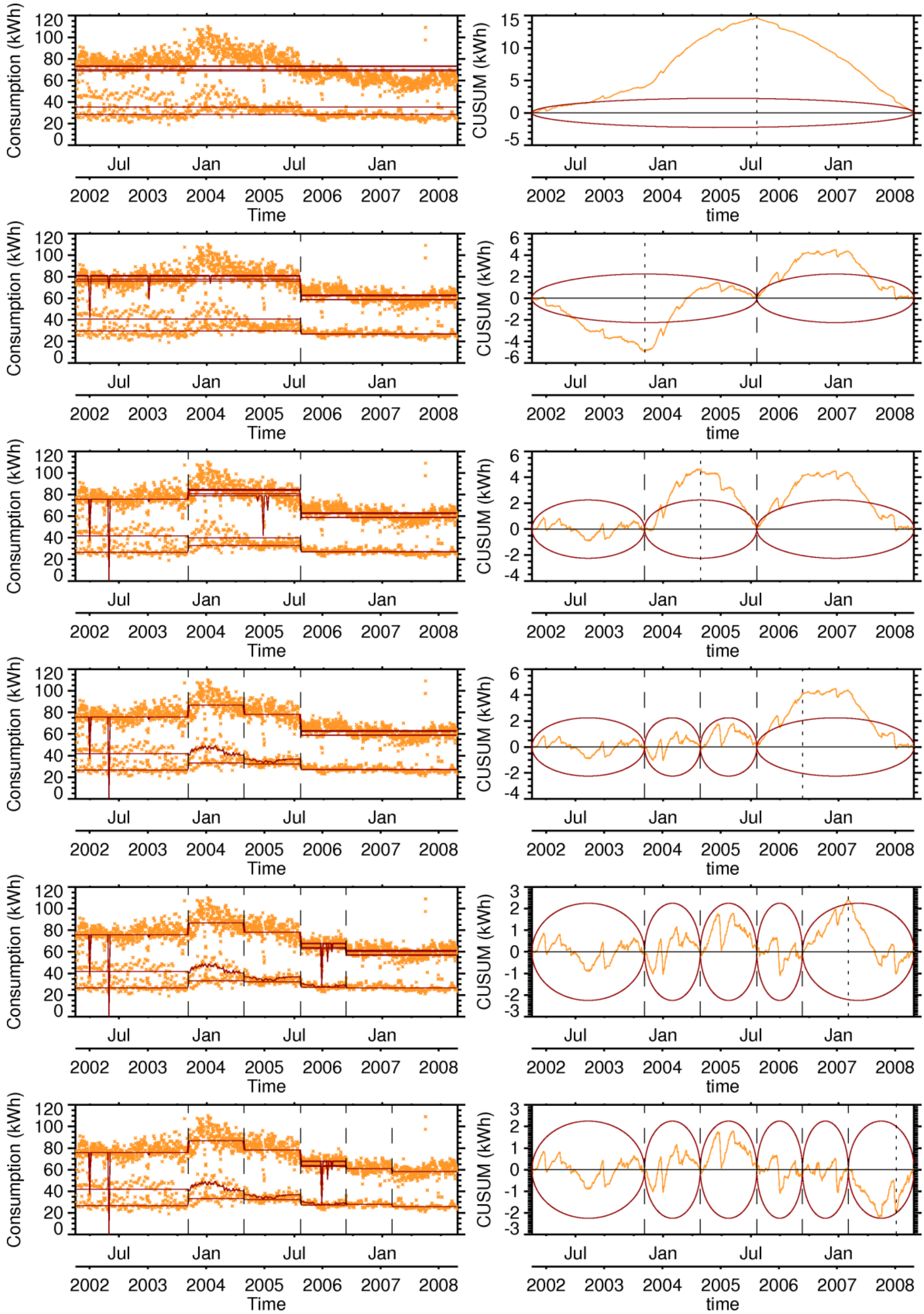


Figure 4.10: Event detection step by step

The figure shows the event detection process step by step. Each row shows modelled consumption and OLS CUSUM with boundaries for the appropriate stage in the process. In this case there are five events and six segments. Each segment is modelled individually. The resultant model describes and quantifies various changes in consumption pattern.

Figure 4.11 shows the final result. According to the model, the dataset has seen one increase in consumption in the autumn of 2003. This is followed by a series of four significant decreases over the next four years. Weekday and weekend consumption have always been significantly different. Consumption levels on Saturdays and Sundays have converged to a low level though they were initially very different. Consumption on Sundays increased between the autumn of 2003 and the autumn of 2005.

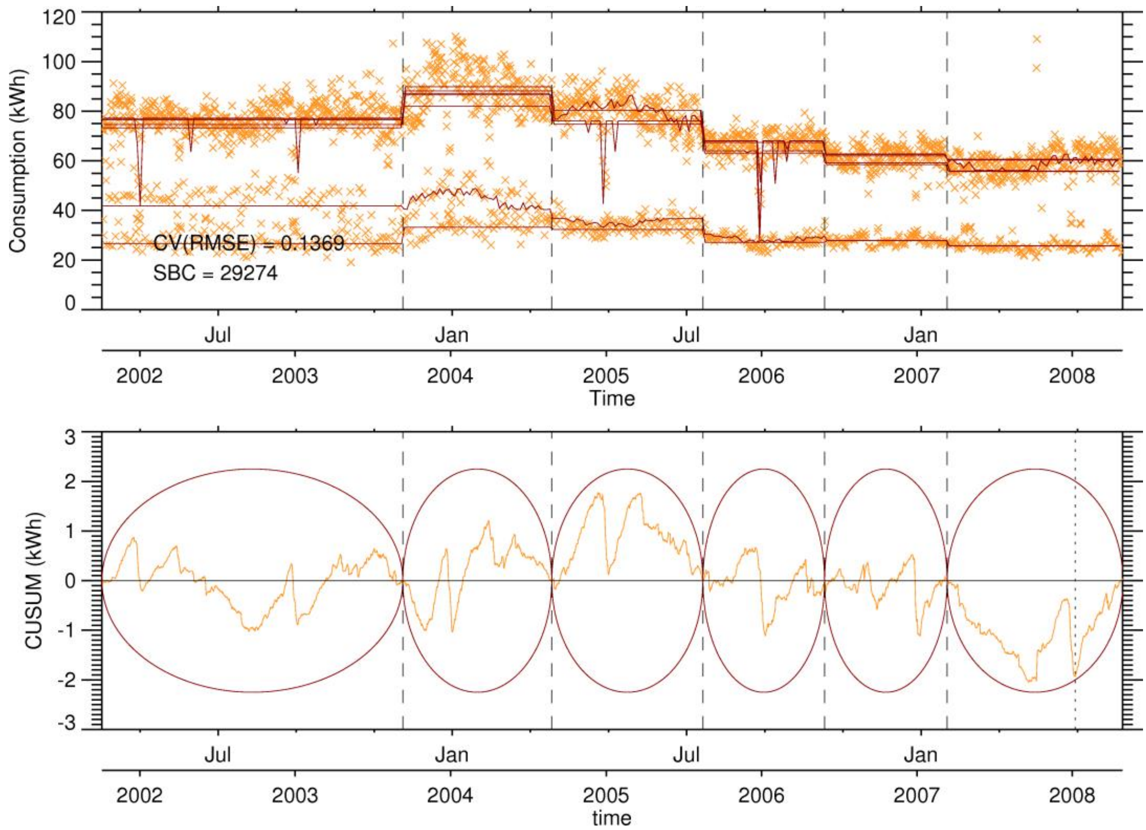


Figure 4.11: Example dataset, model and CUSUM with all events

These insights constitute the output generated by this analysis methodology. They are determined automatically in a few seconds on a moderately powerful PC and can be generated very easily for many datasets. The next chapter presents the results of applying this analysis methodology to the datasets introduced in Chapter 3.

Chapter 5 Results

"However beautiful the strategy, you should occasionally look at the results"

– Winston Churchill (1874 – 1965)

This chapter presents the results of applying the analysis methodology described in Chapter 4 to the data introduced in Chapter 3. The analysis was conducted in seven stages and all results stored in a custom database as described in section 5.1. The analysis was conducted with two specific objectives. Firstly it was intended to demonstrate the impact of the various layers (core model selection; occupancy variation; and event detection) in the methodology. In section 5.2 the impact of each layer of the analysis is demonstrated very clearly in terms of goodness-of-fit metrics and model selection.

It was also intended that a full analysis should be conducted on the entire dataset so as to present the detailed information that this methodology is able to extract. Rather than demonstrate practical uses of the methodology (which is reserved for Chapter 6) the intention in section 5.3 is to introduce the detailed event-oriented meta-data generated by the approach.

5.1 Overview

The analysis methodology as described in Chapter 4 is composed of three simple layers. This section reviews these layers and introduces an experimental design whereby seven different analyses are conducted to demonstrate the impact of each layer in turn. The seven analyses are listed in Table 5.1. Each analysis is given an identifying analysis code which will be used throughout this chapter.

Table 5.1: Analysis parameters for seven experimental runs

Analysis code	Core model Type	Occupancy pattern	Event detection
A	Type 1	[0000000]	None
B	Type 2	[0000000]	None
C	'smart'	[0000000]	None
D	'smart'	[0123456]	None
E	'smart'	'smart'	None
F	'smart'	'smart'	$\alpha << 0.001$
G	'smart'	'smart'	$\alpha = 0.001$

5.1.1 Core model selection

The first layer is the choice of the underlying model. The underlying model can be fixed as either the type 1 (constant) model, the type 2 (VBDD) model or can use 'smart' model selection that will pick the most appropriate model (type 1 or type 2) based on an inspection of the Schwartz Bayesian Criterion (SBC) of the data fitted to each model.

The second column in Table 5.1 represents the choice of core model. In analyses A and B the core model is fixed to type 1 and type 2 respectively. In all subsequent analyses the 'smart' model is employed and the core model is fitted according to the SBC comparison.

Analyses A and B are intended to be compared to highlight the effect of the core model and can also be compared with analysis C to highlight the effect of using SBC to select the core model.

5.1.2 Occupancy pattern selection

The second layer in the methodology is the choice of occupancy pattern. The occupancy pattern is applied by splitting the data into the appropriate subsets (e.g weekdays and weekends) and fitting the core model to each subset in turn.

For brevity, each model variant will be labelled with a series of seven integers. This reflects the IDL implementation of occupancy variation and provides a concise means to describe the different model variants. Each variant has a unique code that describes it fully.

The position of each of the seven integers represents a day of the week. The first represents Sunday and the last represents Saturday. The value of the integer represents the sub-model applied to that subset. That is, the number of unique integers in a given code represents the number of actual sets of model parameters calculated and the number of subsets the data were split into.

For example, when the base variant is fitted there is only one sub-model (model zero) which is fitted to all the available data. The data are not split at all. The base variant is represented by [0000000] because each of the seven days of the week is fitted to this model zero.

Contrast this with the weekday/weekend model which requires two models (model zero and model one). The equivalent label is [0111110] because the data are split into two subsets; the first consisting of all weekend days (represented by zeros) and the second, larger subset consisting of all week days (represented by ones). Model zero is fitted to the first subset and model one is fitted to the second subset.

The full weekly variant [0123456] is the most complex model variant. There are seven subsets and seven models, one for each day of the week. A different model is fitted to each subset in turn and seven different sets of model parameters are generated.

The ‘smart’ variant compares four common alternatives and picks the one which generates the lowest value for SBC. The four alternatives compared by the ‘smart’ variant are [0000000], [0123456], [0111110] and [0111112]. The ‘smart’ variant will always fit one of these four alternatives.

The third column in Table 5.1 represents the choice of occupancy pattern. In analyses C and D the occupancy pattern is fixed to [0000000] and [0123456] respectively. In all subsequent analyses the ‘smart’ variant is employed and the occupancy pattern is chosen according to the SBC calculation.

Analyses C and D are intended to be compared to highlight the effect of adding occupancy variation. They can also be compared with analysis E to highlight the effect of using SBC to select the model variant.

5.1.3 Event detection

The third and final layer in the methodology is the application of the event detection algorithm. Event detection is applied as a binary recursion to split datasets into periods of consistent consumption pattern. It relies on fitting a model to the data and inspecting the stability of that model over time. The model applied to each period resulting from this splitting process is determined by the first two layers.

Event detection is also controlled by the selection of the significance value, alpha. Different values of alpha require the use of different critical values as shown in Table 2.3 in section 2.5. Lower values of alpha will produce fewer false positives but will also produce fewer events. Higher values of alpha will be more sensitive but will produce more false positives.

The fourth column in Table 5.1 represents the use of event detection and the choice of alpha. Event detection is only used in analyses F and G. In analysis G alpha is set to 0.001 or 0.1%. The critical value used is 4.5, this is the least sensitive value provide by Zeileis as replicated in Table 2.3 and should provide one false positive in 1000 events. In analysis F a less sensitive analysis was required but no critical values were provided to go beyond an alpha of 0.001. So, a very large critical value of 12.0 was chosen for comparison.

Analyses E, F and G can be compared directly. Analysis E includes no event detection; analysis F includes event detection whereby only highly significant events are detected; and analysis G includes event detection using a more sensitive significance setting. It is certain that analysis F will produce fewer events than analysis G.

5.1.4 Data management

Each dataset under analysis was analysed seven times according to the definitions in Table 5.1. In total, 738 datasets were analysed (26 datasets were excluded due to too much missing data). For each analysis this key information was recorded in a database along with the defining configuration of the analysis and some basic descriptive statistics. The database schema is shown in Figure 5.1.

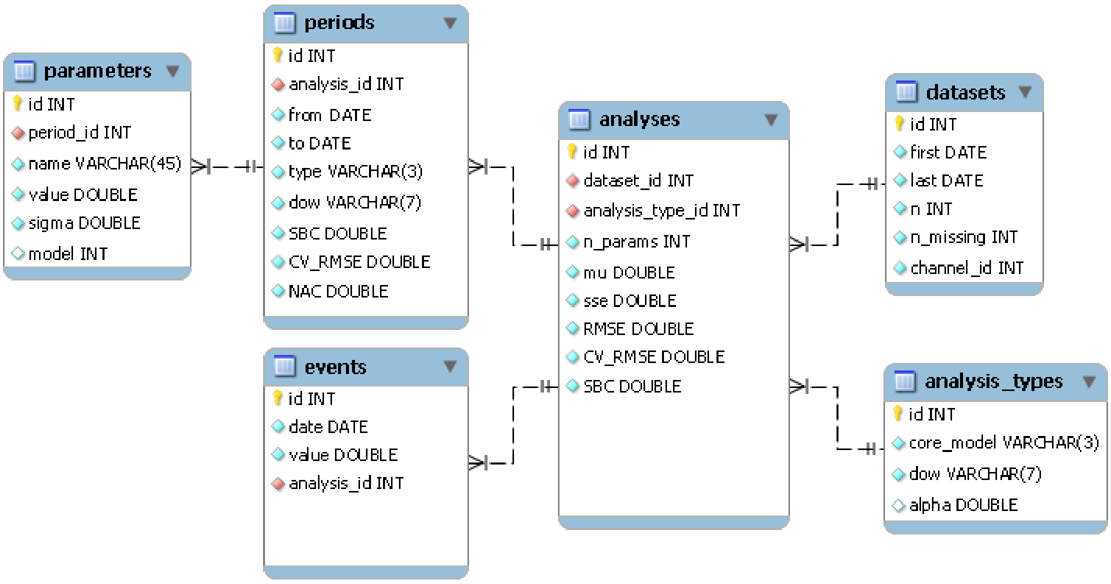


Figure 5.1: database schema for storing results

For each table there is a field for a unique identifier (or ‘primary key’) which can be used as a reference (or ‘foreign key’) in other tables. This links the information in each table so a record in one table can be said to include information from another table. In this way an analysis is linked to both a dataset and an analysis type. Similarly, events and periods are linked to an analysis and parameters are linked to a period. These connections can be seen on Figure 5.1.

The `analysis_types` table holds the definition of each analysis conducted in this work. An analysis definition consists of a core model type, a day of week occupancy setting and, if event detection is employed, a value for alpha.

Each dataset was analysed in turn. For each dataset a record was inserted into the `datasets` table with basic statistics such as the number of data points, the number of missing data points and the period covered by the dataset. This dataset information is linked to the source database via the `channel_id` field.

For each analysis the primary results were recorded in the `analyses` table. These records include the coefficient of variation of the RMSE, the Schwartz Bayesian Criterion and a count of the total number of parameters in the model.

Where event detection was used, the `events` and `periods` tables were also populated. The date and significance of each event was recorded as a record in the `events` table. Information such as the normalised annual consumption (NAC) was also calculated for each period and recorded in the `periods` table.

A period record represents a consumption model and estimated parameters. For each period record generated the individual model parameters were recorded in the `parameters` table. Parameters were recorded along with information to identify which day of the week the parameter applies to and which kind of parameter it is.

The product of this approach is a large database of results including 738 dataset records, 5,166 analysis records, 5,777 event records, 10,943 period records and 54,785 parameter records. These results represent a detailed description of the changing patterns of consumption in the original data.

5.2 Model fitting

In this section each of the seven analyses are described and compared. This discussion will centre on the summary meta-data in the `analyses` table, in particular the goodness of fit and model selection will be considered. It will be shown that, as the modelling process becomes more sophisticated, the goodness of fit increases. It will also be shown that the 'smart' model selection has a negligible effect on model fit whilst ensuring the simpler model variants are employed wherever possible.

Each event detected not only adds a parameter for the event itself but also adds a sub-model with a full set of parameters. This means that a greater reduction in modelling error is required if event detection is to be favoured by SBC. Thus we might expect SBC to select against using event detection in some cases.

Table 5.2 shows some summary statistics for ε for each of the seven analyses. Each row in the table represents a summary of the analysis of all 738 datasets. The table includes the analysis code, the mean and median values of ε, the number of datasets for which the ε is greater than 200% (see section 5.2.1 below) and the number of datasets for which this analysis type produces the lowest SBC.

Table 5.2: Summary of ε for seven analyses

Analysis code	Mean ε	Median ε	> 200%	Count
A	101.2%	69.4%	35	10
B	90.4%	56.5%	32	5
C	90.6%	56.5%	32	0
D	80.7%	43.3%	31	15
E	81.0%	43.4%	32	81
F	70.4%	39.2%	24	45
G	60.9%	31.6%	19	582

The average model fit as quantified by the mean ε improves (decreases) as more and more complexity is added to the model. Also, as indicated by the falling median ε, more datasets are being modelled with less error. However, each individual dataset has a different response. The remainder of this section presents the results in more detail.

The SBC statistics themselves are not of much interest and are not presented here. It is more interesting to look at which model was chosen for each of the 738 datasets. The final column (labelled 'Count') in Table 5.2 shows the number of datasets which, when analysed with that modelling regime, produced the lowest value for SBC.

Some analyses are exactly equivalent. For example, analysis C is always the same as either analysis A or analysis B because it simply chooses between them using SBC. This means that some analyses will produce exactly the same SBC for a given dataset. To produce a count which avoids double counting, the simplest model is always selected when two or more analyses produce the same SBC.

Thus, for analysis G to be selected it must produce an SBC which is ***less than*** that of all other analyses. For analysis A to be selected it need only produce an SBC which is less ***than or equal to*** that of all other analyses. For analysis D to be selected it must produce an SBC which is ***less than*** analyses A, B and C but ***less than or equal to*** analyses E, F and G.

Analysis G, with sensitive event detection and fully automatic model selection, was selected for most datasets. Simpler analyses were selected for some datasets but this is primarily due to the more complex 'smart' analyses producing models exactly equivalent to the simpler cases.

For example, if events were detected in neither analyses F nor G then analyses E, F and G would produce identical results. In such a case, analysis E would be selected unless the smart occupancy detection fitted either the [0123456] or [0000000] model variants; in which case the results would also be identical to that produced for analysis D or C respectively.

If the [0000000] model variant was fitted by analysis E then the results would be identical to analysis C and they would also necessarily match either analysis A or B depending on the core model selected. This is why analysis C is never selected, because it is always equivalent to either analysis A or analysis B.

5.2.1 Core model selection

The first three analysis codes in Table 5.1 refer to the simplest case where there is no event detection and no occupancy variation. The only difference between them is the choice of core model. Analysis A uses the type 1 model, analysis B uses the type 2 model and analysis C uses the SBC to determine which model to apply.

Figure 5.2 shows three histograms, one each for analyses A, B and C from Table 5.1. Each histogram presents the frequency distribution of ε over all 738 datasets. The histograms are truncated at 200% since only a small number of datasets (as shown in Table 5.2) exhibit values of ε higher than this and they are spread over a very large range.

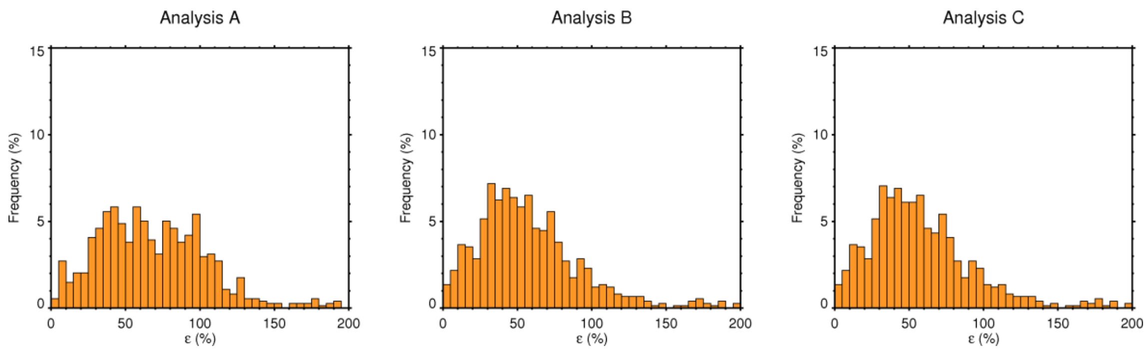


Figure 5.2: Goodness of fit histograms for core model selection

A low value of ε represents a better fit to the data than a high value. In general, all three histograms show a peak in ε which tails off at around 150%. This means there are very few datasets which produce ε greater than 150%. The histograms all show peaks between around 50% and 100%. Few datasets are modelled with a value of ε of 10% or less.

The figure shows that a greater proportion of datasets are on the left (better fit) side of the distribution (i.e. ε < 100%) in analysis B than in analysis A. This shows that the more complex model produces a better fit overall and is to be expected.

Analysis C uses the SBC to determine model type. the more complex type 2 model was chosen over the simple type 1 model in 576 of the 738 datasets. In 162 individual datasets the model fit did not improve very much when moving from the type 1 to the type 2 model. These datasets, where the improvement in fit is negligible, are better fitted to the type 1 model.

This makes the results for analysis C a mixture of the two models with some datasets being fitted to the type 1 model and some (only those where the model fit is significantly better) being fitted to the more complex type 2 model. Figure 5.2 shows that the move to analysis C has a minimal impact on model fit.

The distributions for analyses B and C are very similar even though analysis C includes 162 datasets modelled with the simple type 1 model. Using SBC appears to provide a good compromise between goodness of fit and model complexity.

5.2.2 Adding occupancy variation

Analysis codes C, D and E in Table 5.1 refer to cases where there is no event detection but the occupancy pattern is varied. In each case the core model is allowed to vary based on SBC as in analysis C. The base case [0000000] is compared with full weekly variation [0123456] and 'smart' variation.

Figure 5.3 shows CV(RMSE) histograms for analyses C, D and E. It can be seen that the addition of full weekly occupancy variation (compare analysis C with analysis D) produces a significant improvement in model fit. This is to be expected due to the increased complexity of the full weekly variant.

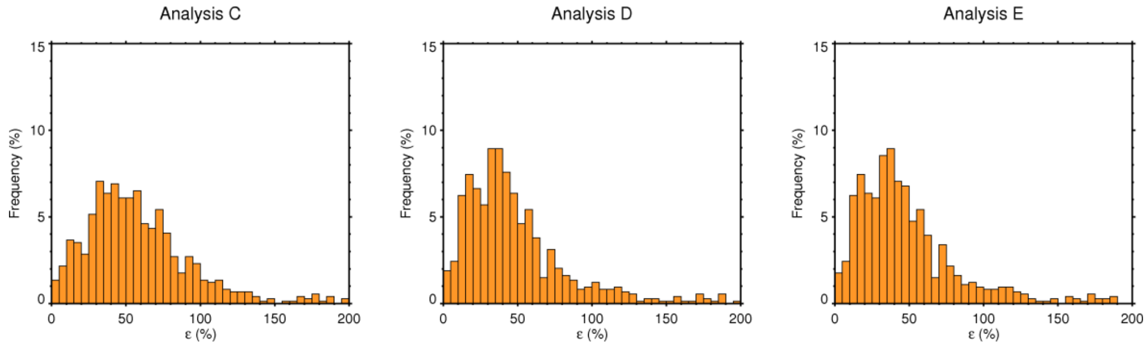


Figure 5.3: Goodness of fit histograms for the addition of weekly variation

The distribution for analysis E is very similar to that of analysis D. This is for much the same reason that analyses B and C were seen to be so similar. The SBC will only choose a simpler model (i.e. a higher ε) if the impact on the goodness of fit is smaller than the penalty imposed for the added model parameters.

For datasets where there is little or no weekly variation the 'smart' variant will tend to favour the simpler model. The SBC for analysis C was lower than that of analysis D for a total of 245 of the 738 datasets. However, an intermediately complex model such as the [0111110] variant may be preferred over either the [0000000] or [0123456] variants.

5.2.3 Occupancy pattern selection

Analysis E allows the weekly pattern to take on one of four alternative occupancy patterns. The alternative chosen by SBC is that which provides the best fit using the fewest model parameters. An increase in complexity must deliver a commensurate improvement in model fit. Thus, given that the alternative patterns are appropriate and provide enough variety to represent most examples, the SBC-selected model is likely to be a close match to the actual consumption pattern.

Table 5.3 shows the number of datasets for which each alternative weekly pattern was selected under analysis E. The equivalent statistics are also shown for analyses F and G, though these represent periods of consumption rather than datasets and so the numbers are larger. For the moment the data of interest is that for analysis E.

Table 5.3: Weekly variation chosen by SBC with 'smart' core model

Weekly pattern	Analysis E		Analysis F		Analysis G	
	Count	%	Count	%	Count	%
[0000000]	163	22.1	393	30.9	2117	35.2
[0111110]	426	57.7	632	49.7	2478	41.3
[0111112]	72	9.8	115	9.0	545	9.1
[0123456]	77	10.4	132	10.4	867	14.4

The most common pattern selected (426 of 738 datasets) is one where all weekdays are modelled together and Saturdays are the same as Sundays. The next most common pattern was one where all days are treated the same with 163 instances. Only 72 datasets showed the weekday/Sat/Sun pattern. A further 77 datasets exhibited enough variation between weekdays to employ the full weekly variant.

Looking at the selected occupancy variant alone is a simplification of what is actually happening here. The analysis also includes a 'smart' core model which, for each subset defined by the occupancy variant, will employ either the type 1 or type 2 model depending on the SBC of each model fitted to that subset. Each of the four cases shown in Table 5.3 disguises a series of different model forms.

For example, of the 163 datasets modelled with the [0000000] pattern there are two alternatives, the type1 model and the type 2 model. In analysis E the type 1 model was fitted to 21 datasets and the type 2 model was fitted to 142 datasets.

More complex weekly patterns can accommodate more alternative configurations. For example, the [0111110] model can employ either model type for weekends and for weekdays. Thus there are four sub-types of the [0111110] model.

Table 5.4: Model variants selected under analysis E

Weekly pattern	Core model types	Count
0111110	2222222	287
0000000	2222222	142
0111110	1222221	68
0111110	1111111	64
0111112	2222222	45
0123456	2222222	29
0000000	1111111	21
0123456	1111111	12
0111112	1222222	11
0111112	1111111	11
0111110	2111112	7
0123456	2222221	4
0123456	1111211	4
0123456	2111111	3
0123456	1222222	3
0123456	1111112	3
0123456	1112111	2
0111112	2111112	2
0123456	1112211	2
0123456	2112222	1
0123456	2111211	1
0123456	2121222	1
0123456	1222122	1
0123456	2111212	1
0123456	2222122	1
0123456	2222212	1
0123456	2122222	1
0123456	2112111	1
0123456	1222221	1
0123456	1222111	1

Table 5.4 includes a list of all the model types selected under analysis E in order of 'popularity'. The table shows the familiar weekly pattern code which identifies the days of the week to be modelled separately but adds the 'Core model types' column. This shows a similar code composed of 1's and 2's.

The core model types code represents the type of model (type 1 or type 2) selected for each day of the week. It is constrained to follow the weekly pattern imposed by the selected model variant but can take on any combination of the two model types within that pattern. Thus, there are two sub-patterns (2^1) within the [0000000] model variant, four (2^2) for the [0111110] variant, eight (2^3) for the [0111112] variant and 128 (2^7) for the [0123456] variant.

The most popular combination of models is a [0111110] weekly variation where weekdays and weekends are different but both are modelled with a type 2 model. The second most popular combination is a [0000000] variant where all days are described by the same type 2 model parameters. In the third most popular combination the weekly variant is also [0111110] with weekdays modelled with a type 2 model but a type 1 model is fitted to the weekends.

The less popular [0123456] model variant includes many more sub-types. Only twenty of the 128 potential variations of the [0123456] model have been selected for use in the dataset under analysis. Many of these are cases where one weekend day is the same as a weekday or where one weekday is different from the rest. It could be that a more parsimonious model than the full weekly model is possible for some of these datasets but an exhaustive search using SBC would be computationally expensive.

5.2.4 Adding event detection

The final three analyses in Table 5.1 refer to cases where both the core model and the occupancy variant are chosen automatically as in analysis E. In analyses F and G event detection is employed at two significance levels, analysis F is set to be insensitive and only select the most significant events, analysis G will detect events only if they are at a significance of 0.001 (0.1%) or better.

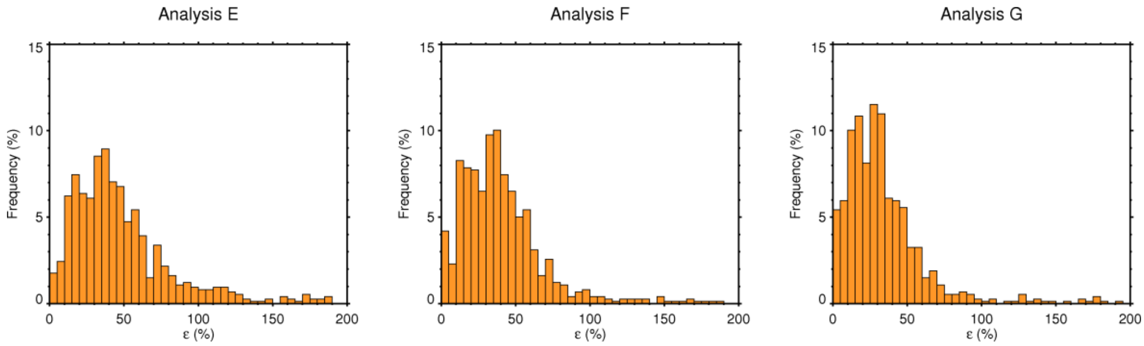


Figure 5.4: Goodness of fit histograms for the addition of event detection

Figure 5.4 shows the now familiar ε histograms for analyses E, F and G. The three histograms show a progression with each successive analysis bunching the histogram up to the left (better fit) side. It is clear that adding the event detection process (comparing analyses E and F) generates a somewhat better fit to the data. With the more sensitive event detection of analysis G, nearly all datasets are modelled within 80% error.

As with the previous examples, this improvement is to be expected wherever a more complex model (in terms of number of parameters) is used. Thus, we should expect the fit to improve when more events are detected. More convincing evidence to support the addition of event detection comes from the SBC model selection.

In Table 5.2 the figures showing which model was 'chosen' by SBC include the rule that, where the SBC is identical the simplest analysis is preferred. If rather than choosing the simpler model, the more complex model is chosen then the figures show a different result. Under these circumstances analysis G is selected for 678 datasets and analysis F is selected for the remaining 60 datasets. There are no cases where the addition of event detection increases the SBC.

The addition of event detection always increases the complexity of the model. The additional model parameters introduced by events always produce enough of an improvement in model fit to justify their addition.

5.3 Event-oriented meta-data

This section describes the `events` and `periods` tables in the results database. These constitute the event-oriented meta-data. Analyses A to E did not include event detection so this section is only concerned with analyses F and G. Results from these two analyses will be presented in parallel for comparison. Further analysis is provided in Chapter 6.

Analysis F detected a total of 534 events, analysis G detected 5,269 events. Since the number of periods in a dataset is always equal to the number of events plus one, the equivalent total number of periods detected for each analysis was the same as the number of events plus the number of datasets, i.e. 534 + 738 = 1,272 and 5,269 + 738 = 6,007 respectively.

The event detection procedure generates two pieces of information about each event. These are the date at which the event was detected and the 'significance' of the event. These were recorded for each event detected in the analyses.

Each period is recorded with information about the start and end points plus the model types and the model fit and the normalised annual consumption for the period. Finally, the actual model parameters are recorded for each period. The following sections describe these data.

5.3.1 Event timing

In the data under analysis the number of datasets increases over time. This was shown in Figure 3.3 in Chapter 3. The number of events detected in each 28-day period also increased over time as shown in Figure 5.5.

The pattern shows a gradual increase in the number of events in line with the increasing number of datasets under analysis. It might be reasonable to expect the rate of occurrence of events over time to be relatively constant. This idea is explored in section 6.1.1.

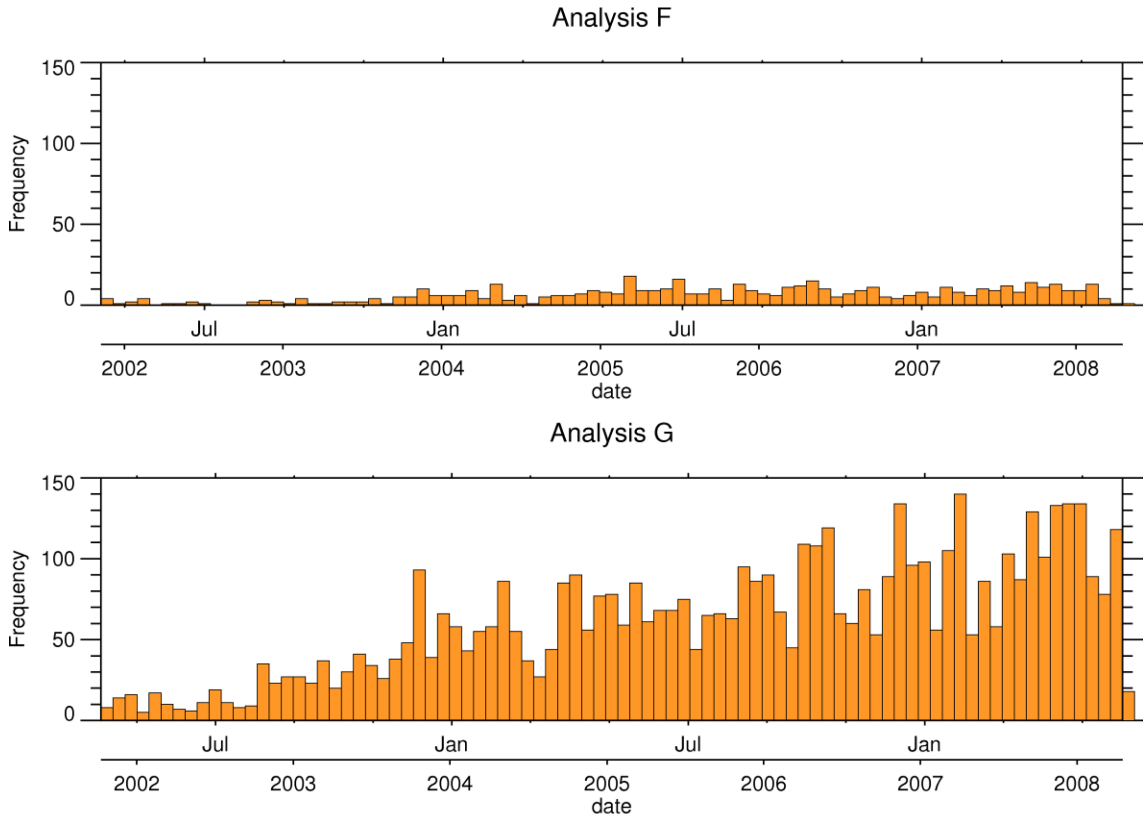


Figure 5.5: Event dates distributions for analyses F and G

5.3.2 Event significance

Event significance was calculated and stored for each event detected. Significance is calculated as λ from equation 4.7. This is effectively how many standard deviations away from zero the OLS CUSUM statistic reached when the event was detected.

The distribution of λ across all events detected in the entire dataset is shown in Figure 5.6. The distribution is broadly symmetrical around zero. Positive values of λ indicate events which caused consumption to reduce; negative values of λ indicate events which increased consumption. The central gap around zero reflects the cut-off point for event detection. In the case of analysis F, the cut-off point was 12.0, for analysis G the cut-off point was 4.5.

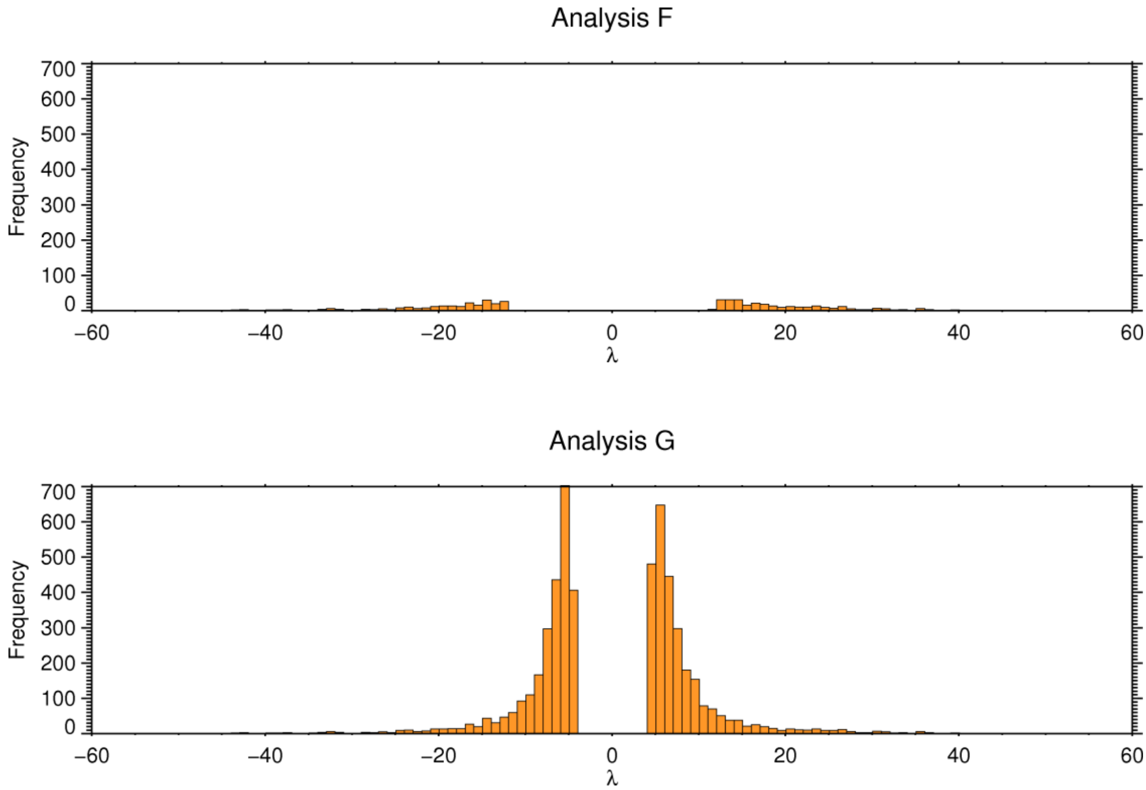


Figure 5.6: Event significance (λ) distributions for analyses F and G

The distribution tails off towards larger absolute values of λ indicating more significant events are rarer than less significant events. Significance is not the same as impact. For an event to be very highly significant its impact must be large relative to model goodness of fit. The impact of events will be explored later on in section 6.1.2.

To clarify the meaning of λ it is useful to consider an example dataset. For this, the dataset with the highest average event significance of all those analysed was selected. Figure 5.7 shows the event detection process for this dataset. The dataset is split by eight events into nine periods. It is clear that there have been a number of major shifts in consumption patterns.

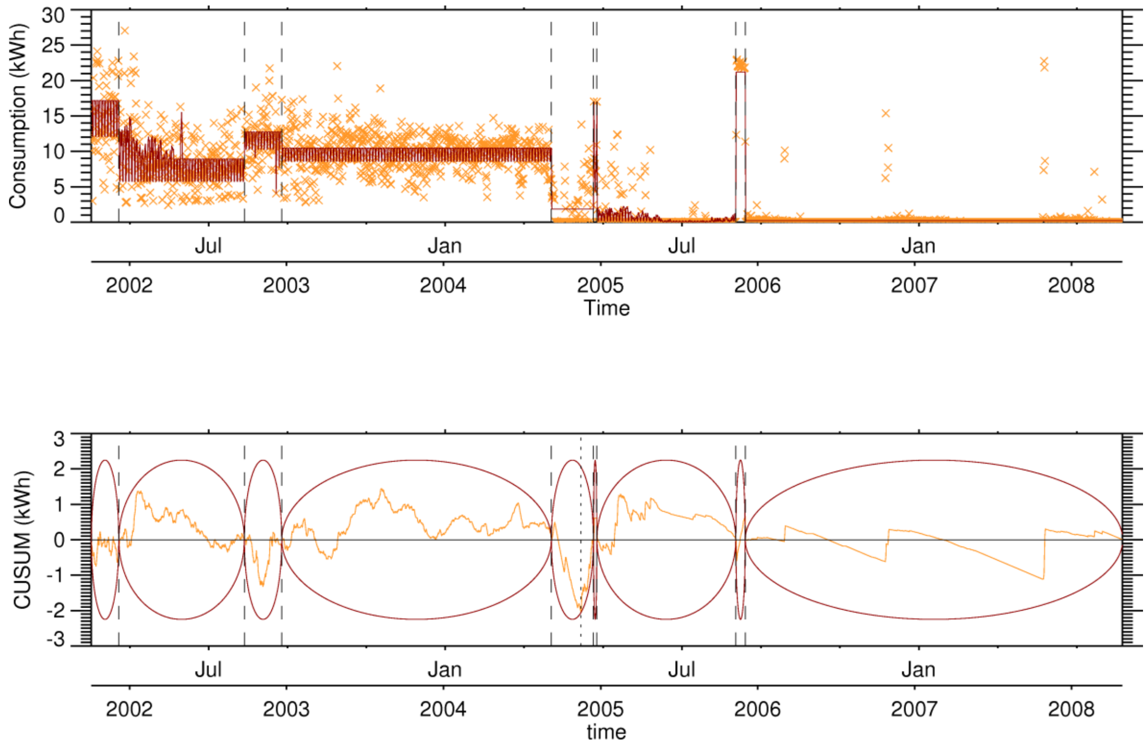


Figure 5.7 : Example of a dataset with highly significant events

By looking at the dataset it is clear that the most important event in terms of energy management was the event in the autumn of 2004. This event represents a persistent move from high consumption to low consumption. Before this event the mean consumption was fluctuating around 10kWh day^{-1}; after the event the mean consumption dropped to less than 1kWh day^{-1}. All other events are relatively insignificant in comparison.

The events are listed with their calculated significance values in Table 5.5. An increase in consumption causes an event with negative significance and vice versa. Looking at the events in the figure and comparing them with the significance values, it is apparent that significance is related both to the level of change in consumption and the order in which the events were detected.

The most significant event is indeed the event identified above with a significance of 38.96. Comparing this value with Figure 5.6 identifies it as one of the most significant events in the whole dataset. This is because the consumption patterns were relatively closely modelled before and after the event, even when it was detected as a single break.

Table 5.5: Events significance values for example dataset

Date	Significance
07/12/2001	11.89
25/09/2002	-7.84
21/12/2002	6.86
08/09/2004	38.96
15/12/2004	-5.98
23/12/2004	6.29
11/11/2005	-16.50
03/12/2005	9.98

This produced a very high OLS CUSUM value and consequently a high value for lambda. This first step in the event detection process for this dataset is shown in Figure 5.8.

All subsequent events were detected relative to either the period before or after this first event and thus they were correctly identified as being less significant. The later four events are two sets of matching pairs, each very close together, perhaps representing short periods of occupancy. The significance of these is smaller than the larger event but still highly significant when compared to Figure 5.6. This is because there is only a short amount of time for the OLS CUSUM to be diverted. This also reveals a quirk in the significance calculation which will be discussed in section 6.2.4.

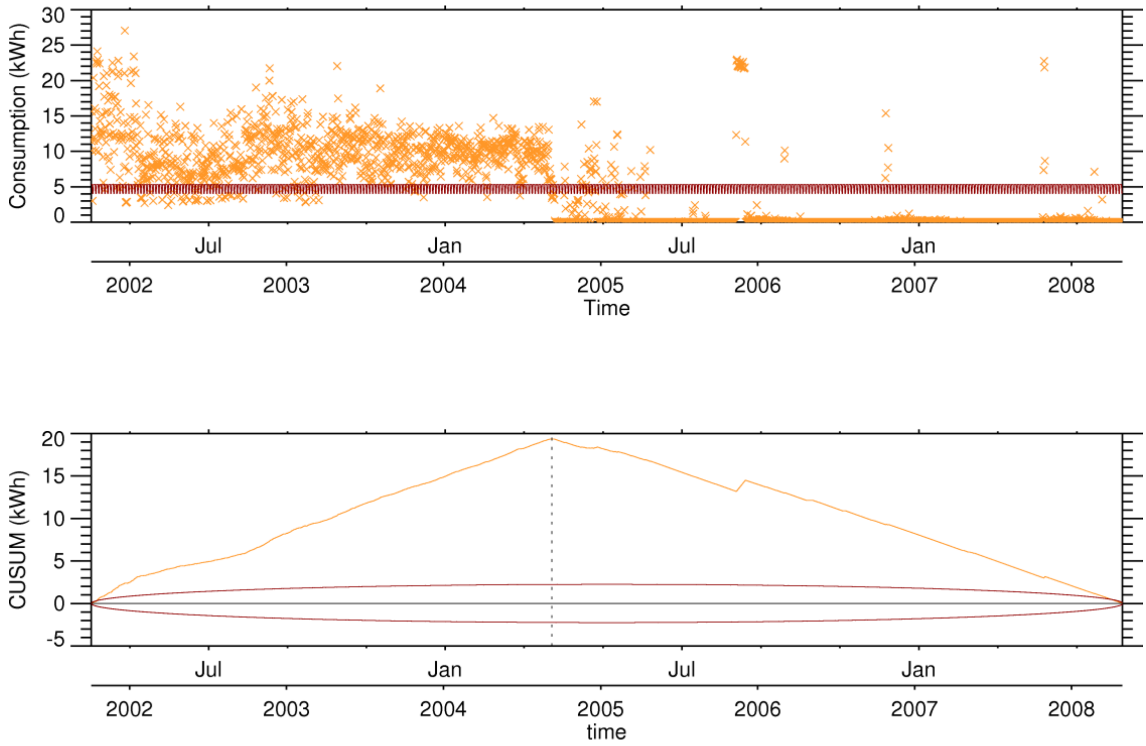


Figure 5.8: Event detection for the first event in the example dataset

5.3.3 Period modelling

The event detection process produces consumption models for each period between events. There are many more instances of each model than for analysis E and they are more reliable than those presented in section 5.2 because they are unaffected by the changes in consumption patterns which lead to events being detected in the first place.

The models chosen to fit each period contain information about the pattern of consumption in each period. This information takes the form of model parameters but also of model types. It was shown in section 5.2.2 that the modelling process leads to a large number of alternative model structures. This is also true in analyses F and G.

The model structures fitted to the data are selected by SBC. There may be a tendency with SBC towards selecting simpler models but it is expected that this process should find the model that most closely matches each dataset. As such, inspecting the selected model types reveals something about the nature of the datasets under analysis.

The top 25 model variants selected under analyses F and G are presented in Table 5.6. The table is sorted according to the number of instances of each model variant fitted under analysis G. Analysis F broadly follows the same pattern though there are a few differences.

Table 5.6: Top 25 model variants selected under analyses F and G

Weekly pattern	Core model types	Count (F)	Count (G)
0000000	2222222	266	1288
0000000	1111111	127	829
0111110	2222222	357	796
0111110	1111111	125	730
0111110	1222221	116	561
0123456	1111111	23	175
0111110	2111112	14	160
0111112	1111111	22	139
0111112	2222222	67	113
0111112	1111112	5	100
0111112	1222222	13	67
0123456	2222222	37	52
0111112	1222221	5	42
0111112	2111112	2	39
0111112	2111111	1	26
0123456	1211111	2	24
0123456	1112111	5	19
0123456	1121111	3	19
0123456	1111121	6	18
0123456	1111211	3	16
0123456	1222221	2	15
0123456	1221111	1	13
0123456	1222222	6	13
0123456	2222221	5	12
0123456	1112211	2	12

As demonstrated in Table 5.3 on page 97, the most common model types are the [0000000] and [0111110] variants, followed by the [0111112] and [0123456] variants. Table 5.6 shows that the type 2 model is generally more popular than the type 1 model and that the most common forms of the [0123456] variant include only one model type followed by forms where six days of the week fit one model and one day fits the other model. In these cases the [0123456] variant may be more complex than necessary.

5.3.4 Parameter count

Another way to explore the detailed model selection process is to use the number of model parameters as a proxy measure for the model complexity. The models in use in the present work have between 1 and 21 parameters. The parameter count for a period is a product of the weekly variation pattern and the model type selected for each of the days of the week.

The [0000000] model variant has one parameter with the type 1 model and has three parameters with the type 2 model. The weekday/weekend [0111110] model variant can have two type 1 models (2 parameters), two type 2 models (6 parameters) or one of each (4 parameters).

Table 5.7: Parameter count for each model variant with 'smart' core model selection

Weekly variation	Parameter count
[0000000]	1, 3
[0111110]	2, 4, 6
[0111112]	3, 5, 7, 9
[0123456]	7, 9, 11, 13, 15, 17, 19, 21

In this way, each variant has its own signature number of parameters and there is little overlap. Table 5.7 shows how the parameter count maps onto the weekly variant. Only periods with 5, 7 and 9 parameters can be interpreted as more than one model type.

A [0111112] variant which uses the type 1 model in all three cases will have three parameters as will a [0000000] variant which employs a type 2 model. A [0111112] variant with two type 2 models and one type 1 model will have seven parameters as will a [0123456] model using the type 1 model for each day. Finally, there are nine parameters when a [0123456] variant uses only one type 2 model and when a [0111112] model uses the type 2 model throughout.

Figure 5.9 shows the distribution of the number of parameters recorded for each period. Most periods are modelled with between 1 and 6 parameters. There are relatively few periods with the more complex models. The [0111112] and [0123456] variants seem to be relatively rare.

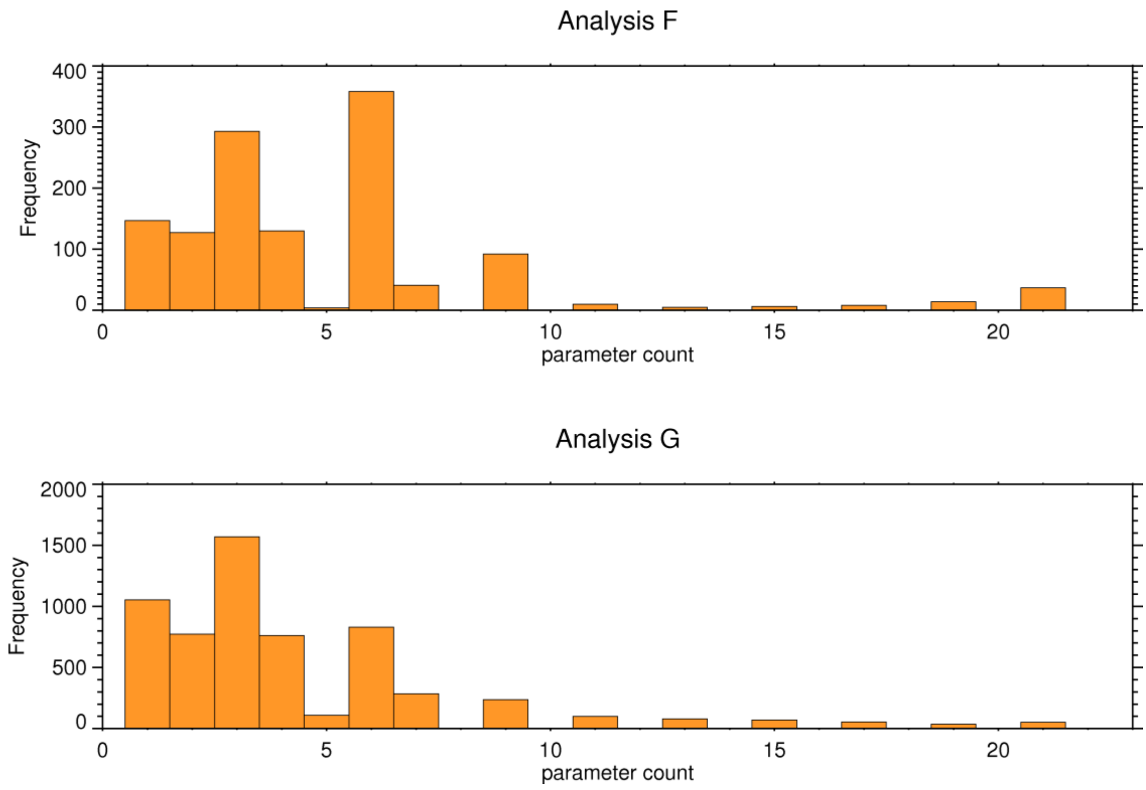


Figure 5.9: Distribution of model parameter count for analyses F and G

It is interesting to see that more complex models are favoured when the less sensitive event detection is used. This can be seen in the relative popularity of the 21-parameter and six-parameter models. This implies that more complex models may be necessary to absorb variation due to simple events. If this is the case then it is clear that event detection improves matters considerably and makes simple models more effective. This would explain the preference for event detection shown by SBC.

Figure 5.10 shows the distribution of the average number of parameters per period for each dataset. This provides, for each dataset a notional idea of the complexity of the pattern of consumption independent of the number of parameters. The distribution for analysis G seems more 'natural' than that for analysis F with a clear peak around four parameters per period.

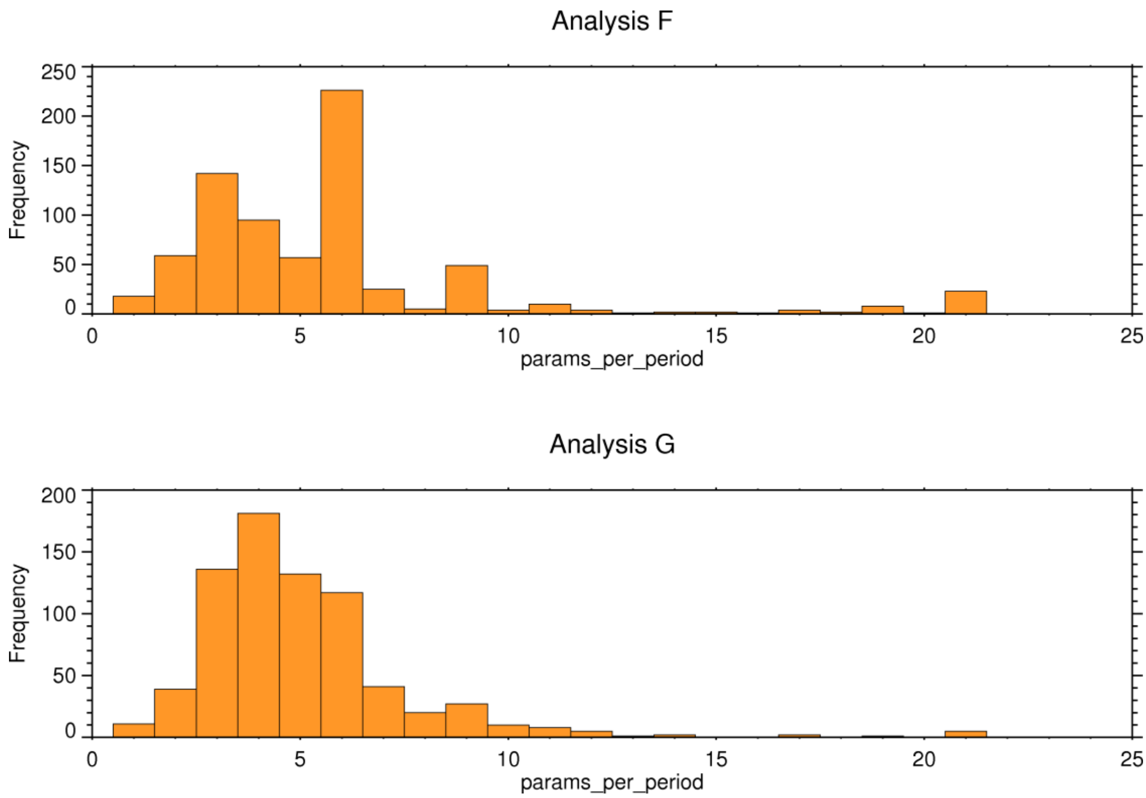


Figure 5.10: Average parameters per period for each dataset (analyses F and G)

5.3.5 Normalised annual consumption

A good measure of the level of consumption in a dataset is the normalised annual consumption (NAC). NAC is calculated as the consumption predicted by a consumption model under fixed conditions, totalled for a year. The calculation determines what consumption 'would have' occurred had the pattern, as defined by the model parameters, been in place during the normalisation period.

NAC was calculated for each modelled period using a common year (the calendar year 2006) of outside air temperature data for normalisation. The distributions of NAC across all periods generated by analyses F and G are shown in Figure 5.11. Most periods have a NAC between zero and 200 MWh/yr, very few have a NAC higher than 500 MWh/yr. The NAC distribution is similar to the average annual consumption.

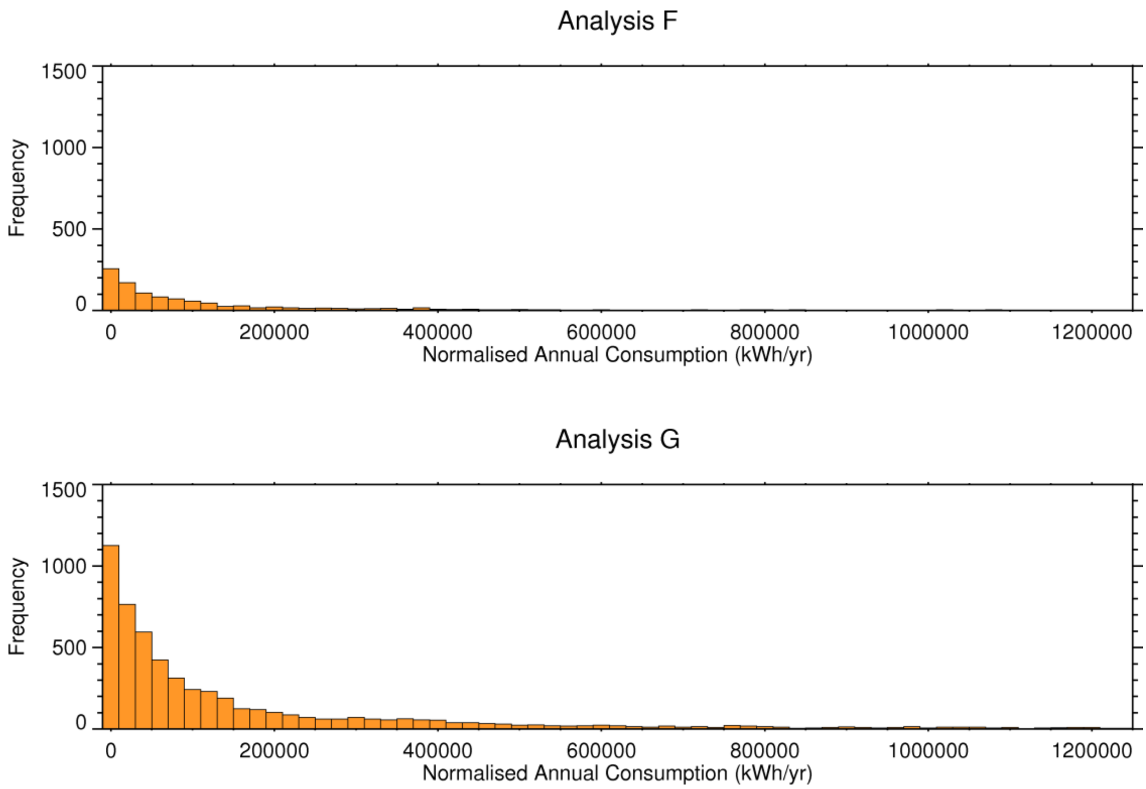


Figure 5.11: Distribution of normalised annual consumption for analyses F and G

NAC figures should be interpreted carefully as the calculation can be performed using less than a year of input data. If a period between events covers only a few months then the temperature range used to determine model parameters is likely to be much smaller than that in the normalisation data which covers a full year. This causes a degree of extrapolation.

It is useful to compare this with the distribution of datasets in terms of their overall level of consumption. Average daily consumption was calculated for each dataset and multiplied by 365 to give a notional measure of average annual consumption. In this case, no account is taken for variations in seasonal consumption. Figure 5.12 shows the distribution of average annual consumption.

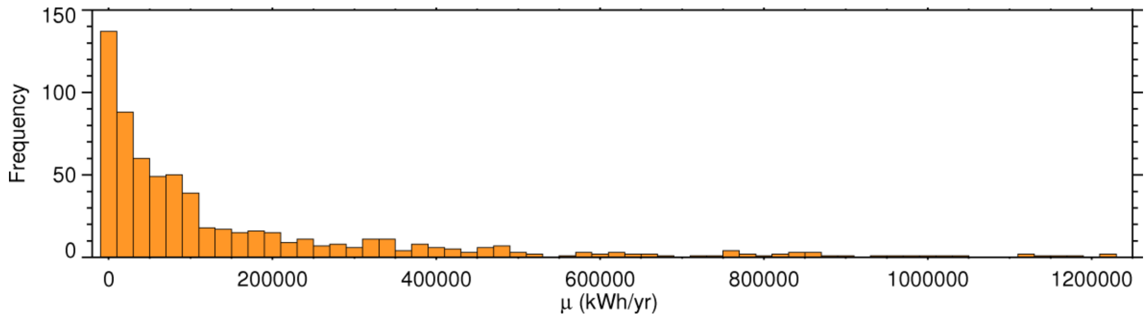


Figure 5.12: Distribution of average annual consumption for all datasets

As might be expected, larger datasets are rarer and smaller datasets more common. There are 81 datasets not shown in the figure because their average annual consumption is greater than the maximum x-axis value.

In Figure 5.11 there are 17 periods where the NAC is lower than the lowest value on the x-axis (i.e. less than zero). This can occur when a period covering a narrow range of temperatures produces an extreme heating coefficient and unusual change-point in the type 2 model.

Figure 5.13 shows the analysis of electricity consumption from a sixth form college building in Leicester; one of only very few examples where the model produces a negative value for NAC. The problem occurs when the type 2 model is fitted to a short period which includes higher consumption during warm weather and lower consumption during milder weather. This can lead to extrapolation with a prediction producing negative consumption during very cold weather which pushes the NAC below zero.

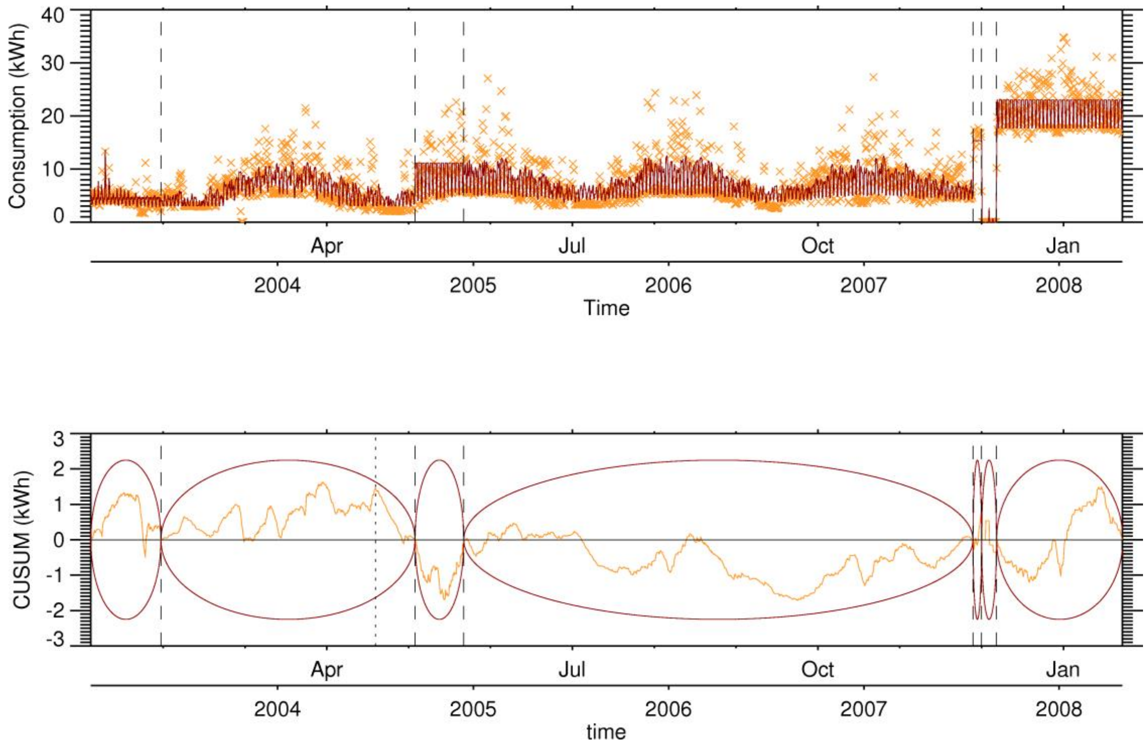


Figure 5.13: Example of period with negative NAC

The period statistics for the analysis are shown in Table 5.8. It can be seen that the model types switch for each period indicating that not only the level of consumption has changed but also the pattern of consumption is changing with each event. The NAC values show that energy performance is generally deteriorating until late 2004 when a stable reduction is achieved. This continues until the autumn of 2007 when consumption levels increase significantly then fall to below zero for a short period before restarting at an increased level nearly three times that before the change.

Table 5.8: NAC values for an example dataset including a negative value

From	To	type	dow	NAC	ε
17/01/2003	28/05/2003	2112211	0123456	1656.8	0.26
28/05/2003	14/09/2004	2222222	0111110	2340.1	0.44
14/09/2004	14/12/2004	2111112	0111110	3391.1	0.38
14/12/2004	24/07/2007	2222222	0111110	2641.6	0.43
24/07/2007	09/08/2007	1111111	0000000	5607.6	0.19
09/08/2007	06/09/2007	1122111	0123456	-515.6	1.96
06/09/2007	28/04/2008	1111111	0111110	7844.0	0.15

In this case consumption apparently fell to zero during the offending period. This was not due to missing data, the system recorded zero values for just over 25 days. Due to the nature of the event detection process and because the metering started recording non-zero values towards the end of a day, the period also includes a single very low but non-zero value. This unusual set of circumstances produced the aberrant set of model parameters. Looking at the values of ε for the periods, it is clear that the offending period is poorly modelled so this kind of problem is usually easy to diagnose.

5.3.6 VBDD Change point

The last database table generated in this analysis is the parameters table. This holds the actual numerical values for each model parameter. There are significant difficulties when attempting to present the values in this table because the values may not demonstrate any particular pattern. The non-heating consumption for example will be largely influenced by the size of the building.

However, The change point temperature in particular is of great interest and is also provided on a simple, common scale. Figure 5.14 shows six distributions, one for each analysis from B to G. Analysis A has no change points as it imposes the type 1 model. Analysis B produces 738 change points, one for each dataset. Analysis C produces 576 change points, one for each dataset for which the type 2 model was selected. Some basic statistics are presented in Table 5.9.

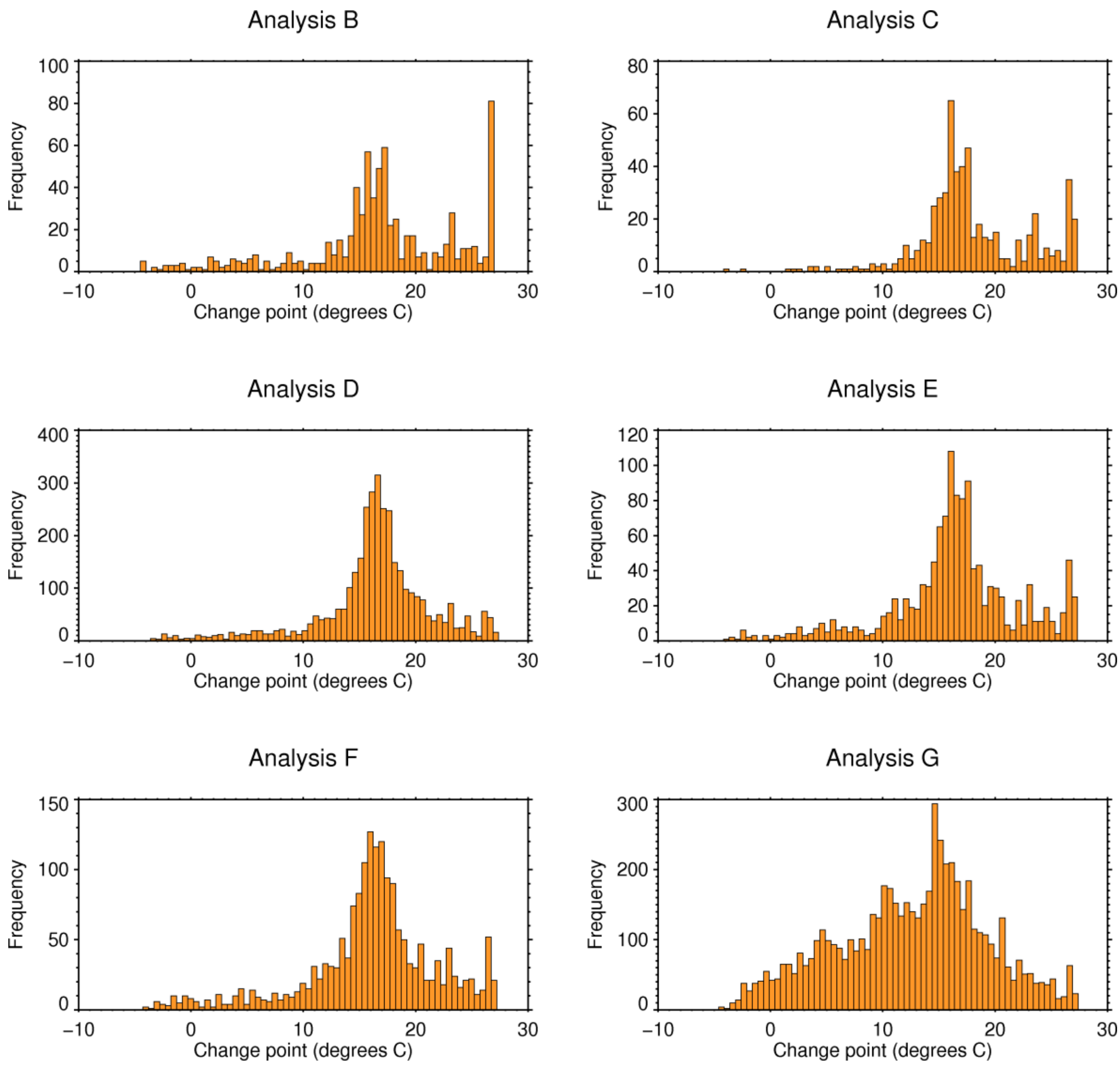


Figure 5.14: Distribution of change point temperature for analyses B to G

The different analyses show differences as is expected. Analysis B includes many datasets which have no relationship with outside air temperature and so are prone to generating extreme change points. These are removed in analysis C.

Analysis D produces many more change points because each day of the week is modelled individually; this produces a smooth distribution as SBC removes those subsets which have no relationship with outside air temperature. The sheer numbers are reduced dramatically in analysis E when the SBC is used to determine occupancy patterns but the shape is largely maintained.

Analyses F and G include events, analysis F produces much the same distribution as D and E but Analysis G departs from the trend. This is because there are very many shorter periods generated by analysis G. This causes both low and high change points to be selected.

Table 5.9: Change point statistics for analyses B to G

Analysis Code	Count	Mean	Median	Mode
B	738	17.00	17.08	26.74
C	576	18.24	17.39	16.08
D	3487	16.64	16.93	16.63
E	1274	16.61	16.78	16.08
F	1817	16.30	16.76	15.97
G	5947	12.78	13.64	14.64

Short periods are modelled in the same way as long ones; the change point is determined by a grid search from the lowest to the highest temperature experienced in the period. However, if the period is in the winter then the highest temperature may be lower than 16.5ºC and will likely be selected as the change point. Similarly, if the period covers the summer, then the change point may be set artificially high in some cases.

The distributions and table show that the change point temperature for most of the datasets under analysis is a degree or so higher than the standard assumption of 15.5. If we accept that 15.5ºC is the appropriate change-point temperature for buildings in the UK then this disparity must be due to either the thermostatic set points being set too high or the internal gains being particularly low or the heat losses being high (See equation 2.16 for the components that influence the change-point). Alternatively, this could be evidence for revising the 15.5ºC standard base temperature.

5.3.7 Example datasets

It appears that there are large numbers of buildings where the base temperature is between 12ºC and 20ºC. Three examples will now be described, one at each end of this range and one in the central peak. These datasets were selected by identifying well-fitting periods with the appropriate value for change point.

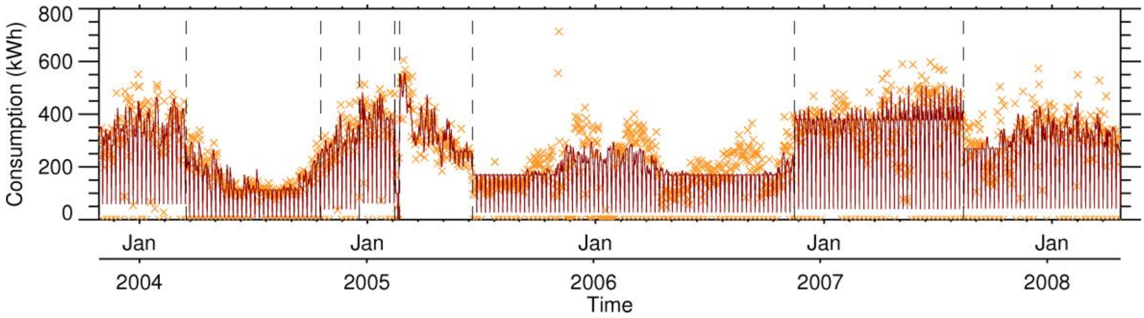


Figure 5.15: Example of dataset with change point of 12.3ºC

Figure 5.15 shows an example dataset where there is a lengthy period (summer 2005 to the end of 2006) during which the change point temperature is estimated to be 12.3ºC. The dataset represents gas consumption in a retail art gallery in Leicester. There are eight events in total separating nine periods. From late 2003 until early 2005 the consumption pattern follows a [0111112] pattern with model types [1222221]. The levels remain fairly constant except for three events relating to changes in the level of consumption on Sundays from 60 kWh day^{-1} down to 6 kWh day^{-1} in March 2004; up to 40 kWh day^{-1} in October 2004; and back up to 61 kWh day^{-1} in December 2004.

After a pair of events in February 2005 the consumption pattern changes to [0111110] with model types [2222222]. It appears the building is heated during the weekend until June of 2005 when the weekend consumption falls back to around 40 kWh day^{-1} and remains there for the rest of the period covered by the dataset. The event in June of 2005 is the beginning of the period with 12.3ºC change point temperature. Prior to this point the change point temperature was 15.2ºC (except for low values during shorter periods only covering the winter). The change occurs in parallel with a significant drop in the heating coefficient from -19.5 kWh ºC^{-1} day^{-1} to -7.7 kWh ºC^{-1} day^{-1}.

The period after the event of November 2006 shows very little control of the heating system with a move to a [0123456] pattern with model types [1121121] and with a low heating coefficient on those days where the type 2 model is applied. Sunday consumption remains at around 40 kWh day^{-1}, consumption during the week becomes fixed at between 343 and 383 kWh day^{-1}. The final event sees a drop in consumption and a return to the type 2 model.

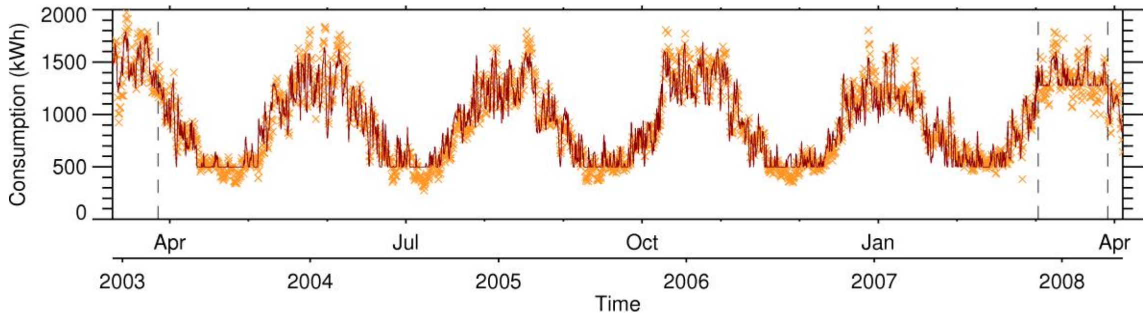


Figure 5.16: Example dataset with change point of 16.7ºC

Figure 5.16 shows an example dataset where there is a lengthy period (spring 2003 through to autumn 2007) during which the change point temperature is estimated to be 16.7ºC. The dataset represents gas consumption in a residential care home for the elderly. Three events were detected in this dataset separating four periods. In all cases the consumption pattern is [0000000] with model type 2. Changes in the model parameters are somewhat misleading as the change point temperature during the three shorter periods is artificially low due to the range of temperatures experienced in those periods. This also impacts the estimate of fixed, non-heating consumption. The only parameter that can be compared is the heating coefficient which begins at -43 kWh $^{o}C^{-1}$ day^{-1}, increases to -56 kWh $^{o}C^{-1}$ day^{-1}, drops back to -46 kWh $^{o}C^{-1}$ day^{-1} and finally falls to -42 kWh $^{o}C^{-1}$ day^{-1}.

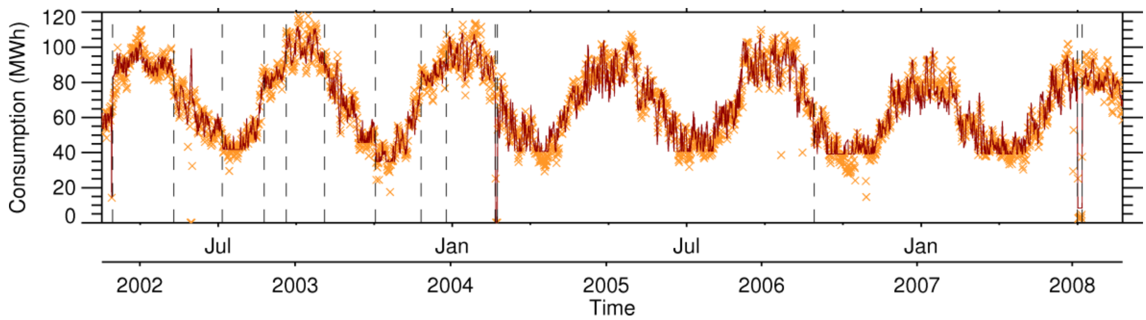


Figure 5.17: Example of dataset with change point of 19.3ºC

Figure 5.17 shows an example dataset where there is a lengthy period (April 2004 to May 2006) during which the change point temperature is estimated to be 19.3ºC. The dataset represents gas consumption in a district heating boiler house. There are a total of 14 events separating 15 periods. All periods follow the [0000000] pattern with model type 2 except for the first period which is fitted to a [0123456] pattern with model types [2112222]. The first nine events have very little impact on model parameters and as such the period before the ninth event is suspected to be affected by ‘phantom’ events as discussed later in section 6.2.4. The period after the ninth event has fixed consumption of 80.8 MWh day^{-1}, a heating coefficient of -2.0 MWh $^{o}C^{-1}$ day^{-1} and a change point of 10.5ºC. The following period is only five days long and so will be ignored. Next comes the period with high change point temperature, it has fixed consumption of 40.6 MWh day^{-1}, a heating coefficient of -2.9 MWh $^{o}C^{-1}$ day^{-1} and a change point of 19.3ºC. The effect these changes have on NAC is to reduce it from 31.8 GWh yr^{-1} to 24.6 GWh yr^{-1}.

Chapter 6 A dynamic model of consumption

"Essentially, all models are wrong, but some are useful."

– George E. P. Box (1919 -)

Chapter 5 introduced a database of event-oriented meta-data which describes the raw metered energy consumption data in terms of simple model parameters changing over time. This database is the output of the analysis described in Chapter 4 is the most significant output provided by the present work.

This chapter demonstrates the value of these data and exposes some difficulties in the method. It is organised in two sections: Section 6.1 explores how useful these meta-data are, i.e. how much they can tell us that was previously unknown; Section 6.2 explores limitations and suggests improvements to the method.

6.1 Analysis

The event-oriented meta-data describe consumption patterns. In this section various approaches to analysis of these data are presented. These methods provide analysis of the meta-data as a whole to give a picture of how the Leicester City Council building portfolio is performing as a whole and how that performance is changing over time.

6.1.1 Period length

Investigating the timing of events is interesting, it might be expected that events occur at random across the building portfolio. This assumption can be tested by looking at the length of the contiguous periods between events. Figure 6.1 shows the distribution (in 28-day bins) of the number of days per period across all periods identified in the dataset.

Period length can be loosely interpreted as the length of time before an event occurs. It should be noted that this analysis is approximate since the calculations include both the first and last periods for each dataset. These periods are artificially cut short as the beginning and end of each dataset does not necessarily represent a change in consumption patterns.

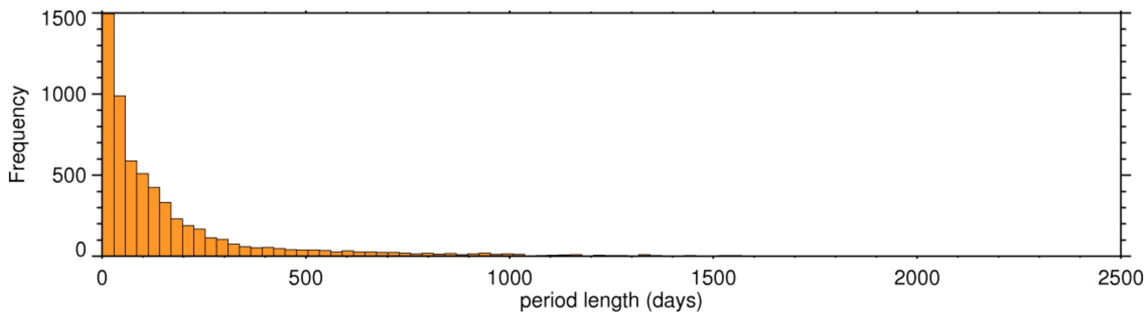


Figure 6.1: Distribution of period length

The distribution is a smooth decrease in frequency as period length increases. It seems period length follows a predictable pattern consistent with a view that events occur at a predictable rate across the whole building portfolio. For each dataset, each day there is a fixed chance that an event will occur. Longer periods are rarer than short periods because they have survived for many days against the odds.

Further analysis can show whether this notional probability of events occurring is the same between different datasets. Figure 6.2 shows the average period length per dataset. It indicates, for each dataset, how long (on average) a consistent pattern of consumption can be expected to continue before an event occurs. This was calculated as the number of days covered by the dataset divided by the number of periods the dataset was divided into.

The distribution is a neat peak mostly spread over the region from one month to one year. There are many examples of datasets where the average period length is longer than one year. Datasets with an average period length longer than 2 years are very rare. There are 20 datasets with average period lengths of 28 days or less.

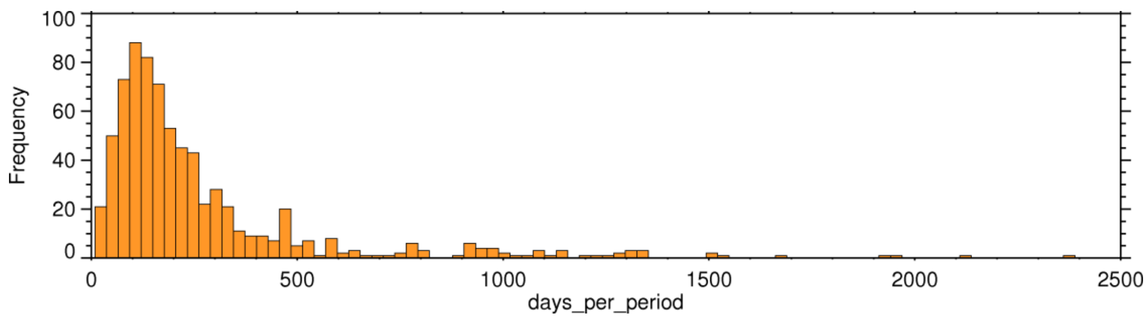


Figure 6.2: Distribution of average period length per dataset

It is clear that, for most of the datasets under analysis, events occur at a rate between 1 $month^{-1}$ and 18 $month^{-1}$. Datasets which maintain longer periods of consistent consumption patterns can be considered more resilient, while other datasets experience events more often. It may be of interest to identify which buildings are consistently able to maintain longer period lengths as these are less affected by events. Long periods may imply that consumption patterns are very tightly controlled and are never allowed to change. Alternative, they may imply that consumption is based on simple, reliable equipment and there are fewer opportunities for a change to occur.

6.1.2 The impact of events

Of greatest interest to the energy manager is the impact of events. It is extremely valuable to know how much effect on overall consumption each event has. With this information it becomes possible to generate a list of the most influential events and to focus most attention and resources on those with the biggest impact. This, together with event significance (see section 5.3.1) is the most compelling justification for an event-oriented approach.

NAC has been calculated for each period in the database (see section 5.3.5) The effect of events can be quantified by comparing the NAC of neighbouring periods. For each event in the results database the change in NAC (ΔNAC) was calculated. That is, the difference between the NAC of the period before the event and the NAC of the period after the event. Figure 6.3 shows the distribution of ΔNAC.

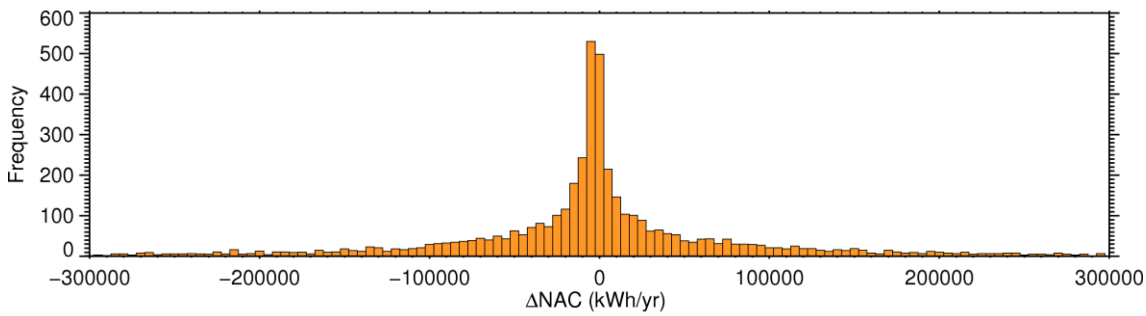


Figure 6.3: Change in normalised annual consumption due to each event

It is clear that the distribution is fairly symmetrical with both negative and positive impacts. Smaller events are more common than larger events. The figure is truncated at -300 MWh yr^{-1} and 300 MWh yr^{-1} to show the distribution of the central peak but there are several events which occur outside this range. There were 387 events lower than the minimum shown and 443 events higher than the maximum shown.

There were 2,587 negative events with a total impact of -6,997 GWh yr^{-1} and 2,677 positive events with a total impact of 6,822 GWh yr^{-1}. Five events had zero impact. Thus, although there were more events which increased consumption than decreased consumption, the total impact of all events was -174.3 GWh yr^{-1}. The mean impact of all events was a reduction of 33.1 MWh yr^{-1} (equivalent to 3.7MW). This reduction was by no means spread evenly across the portfolio.

6.1.3 Normalised impact

Directly comparing the change in NAC between all events in the database is the best way to isolate events which have the biggest impact on energy consumption and emissions. However it is also interesting to normalise the change in NAC by the average annual consumption of the dataset in question. This reveals how important each event is in the context of the dataset it was detected in.

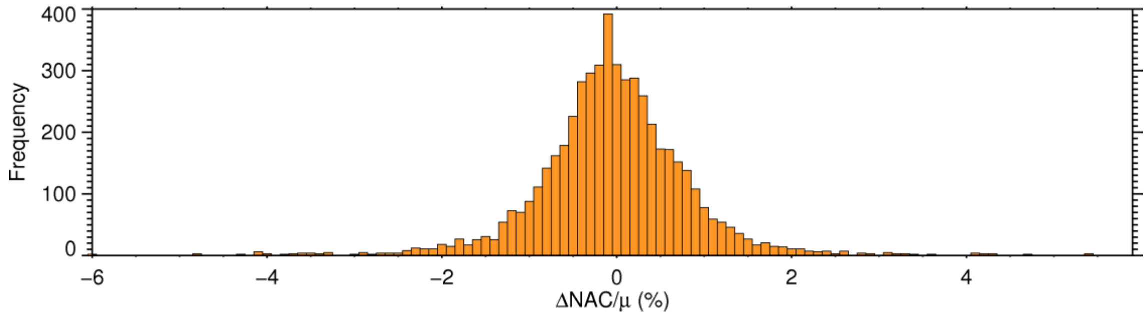


Figure 6.4: Distribution of the normalised impact of events

In Figure 6.4 the change in NAC is presented as a proportion of annual consumption for all events detected. The most obvious feature of the figure is that the histogram follows a normal distribution very closely. The impact of events on average consumption levels rarely exceeds twice the average annual consumption. The number of events with an impact very close to zero seems to be slightly higher than would be expected from a normal distribution.

The symmetry of this dataset should not be interpreted as meaning that every reduction in consumption is matched by an equivalent increase. These figures are normalised so they only show the relative size of each event. A reduction of 200% in one dataset may be matched by an equivalent 200% increase in another dataset but the actual values involved could be very different.

It is interesting to consider whether the symmetry of this dataset implies that there is no bias towards events which reduce consumption against those which increase consumption. It may be assumed that, all things being equal, there is no reason for events to be biased in either direction. It may therefore be expected that the addition of intentional energy efficiency interventions would bias the distribution to the left side.

However, it equally be assumed that the natural state of affairs is a distribution biased to the right with random events, more often than not, causing increases in consumption. Under this assumption, the efforts of energy management at Leicester City Council can be seen as successful in stemming the increase in wasted energy.

6.1.4 Strategic analysis

This section describes two kinds of portfolio-wide analyses that may be of interest from a strategic perspective. The total change in NAC over the period under analysis provides a cross-sectional (dataset by dataset) view of performance to date. A time series of total NAC across the whole portfolio gives an intuitive longitudinal comparison of the most significant events in their proper historical context.

For each dataset, the total impact of all detected events is equivalent to the difference between the latest (i.e. current) period and the first available period (i.e. the base line performance). Viewing this as a distribution across all datasets provides a picture of the overall performance improvement for the entire portfolio.

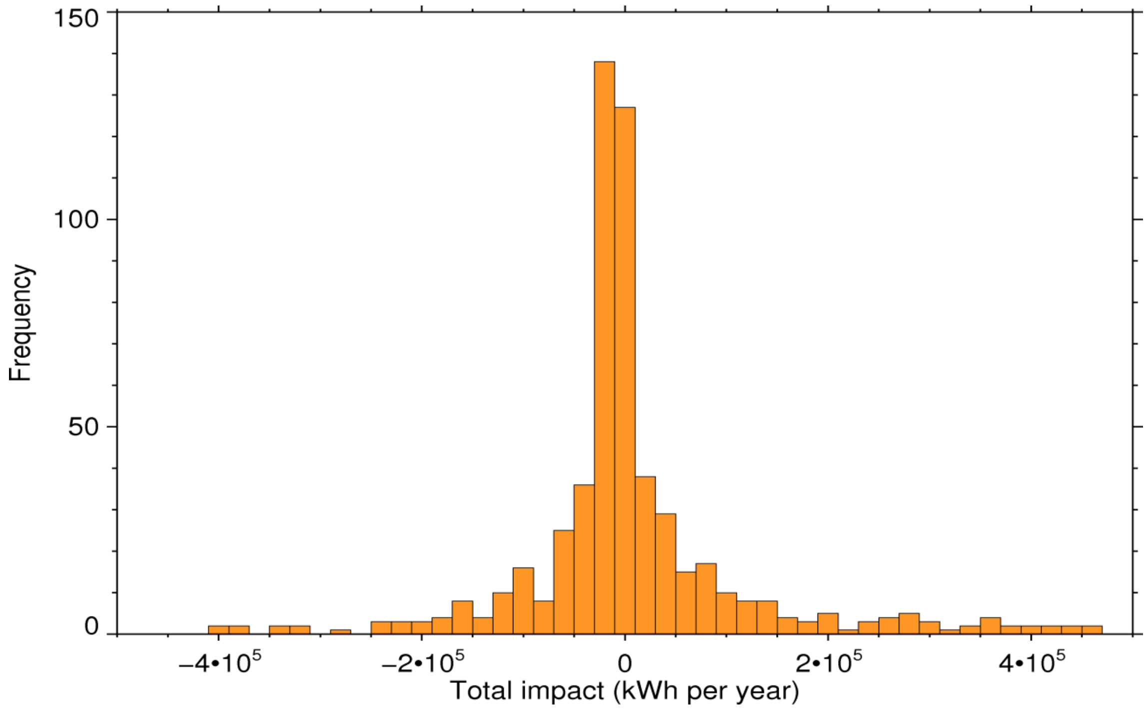


Figure 6.5: Distribution of the total event impact per dataset

Figure 6.5 shows the distribution of the total impact of all events per dataset (in kWh yr^{-1}). The distribution is biased towards the left, indicating improved performance. However, those datasets that contribute to the right hand side of the chart have increased their normalised annual consumption over the period under analysis. We have seen an example of this in Figure 5.13 where the increase in NAC from beginning to end was 6,187.2 kWh yr^{-1}.

Each dataset has a known set of events and between those events the NAC is fixed and known. Thus, it is possible to produce a portfolio total NAC for any given day by summing the NAC for the appropriate subset of periods. Producing a figure for each day over the history of the data is not quite so simple.

One problem with this approach is the effect of new datasets being added into the database over time. If these are not accounted for then the portfolio NAC simply increases over time reflecting the addition of buildings. For a sensible, strategic view of performance it is necessary to ensure the chart only includes data from those datasets which cover the entire period under analysis.

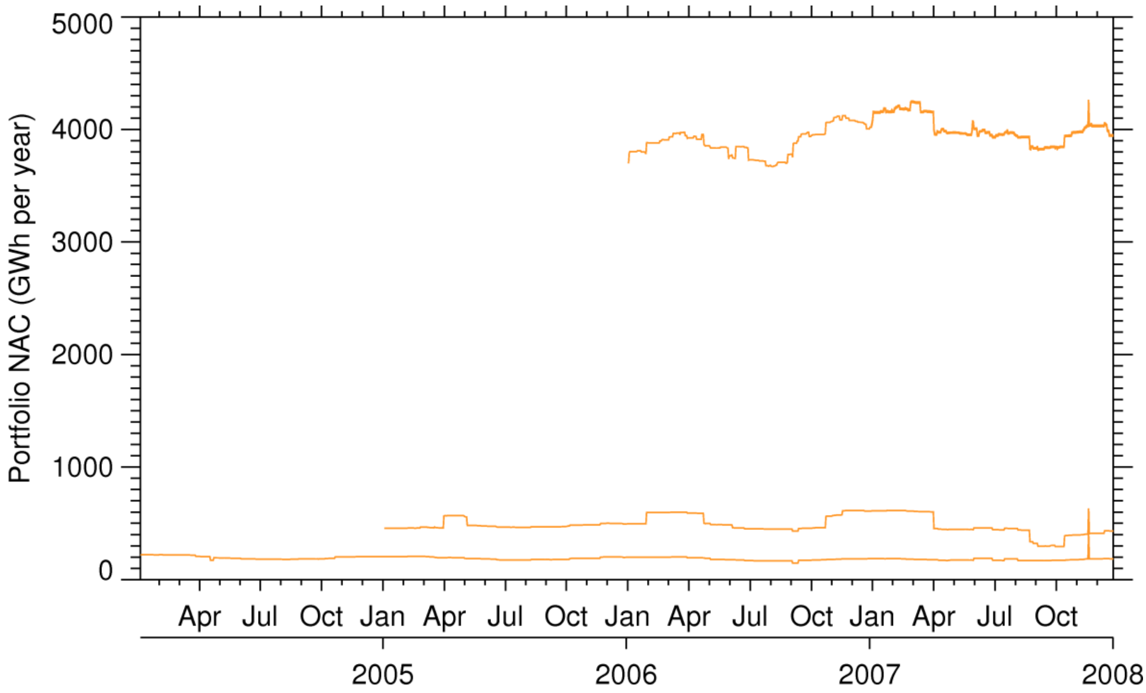


Figure 6.6: Total daily NAC for entire building portfolio

Figure 6.6 shows the results of this calculation for four periods. Each line shows the total NAC for a given subset of the data. The lowest line represents all those datasets which cover the full period between January 1st 2004 and January 1st 2008. The next line up shows the same for the period beginning January 1st 2005. There is a very large jump in NAC when datasets covering the period from January 1st 2006 are plotted. The final line from January 1st 2007 is almost identical to the previous one and can barely be distinguished.

The chart shows the effect of additional datasets as the shift between the different lines. It can be seen that the later plots include the data from the earlier plots. For example, the spike in late 2007 is replicated in all datasets and the large step change in the 2005-2009 portfolio NAC are apparent in the later plots.

Changes in the total NAC over time are the aggregated impact of all events in the database. Because the effects of weather are taken into account, the NAC figures are relatively flat and unaffected by seasonal variation. However, the latest two plots show evidence of a distinct seasonal effect. This is thought to be due to the effect of solar gains. Since seasonal variation in solar gains is not accounted for in the VBDD model it is likely that events could be caused in buildings which are affected by solar gains such that NAC appears to increase in the winter and decrease in the summer.

Figure 6.5 and Figure 6.6 provide a high level strategic summary of the event-oriented meta-data, they provide a means to isolate buildings where performance has deteriorated significantly as well as those where performance has improved significantly. They allow for a global view of when changes have occurred and how big they are. Drilling down into the detailed, rich data underneath this simple view provides a direct link between the strategic and diagnostic data.

6.2 Limitations and improvements

Though buildings may be predictable they are filled with people and so can behave in unexpected ways. There are many imperfections in the method developed and elucidated in this work. A key point to note is that any model of reality is just that; a simplification, a mathematical tool for representing consumption patterns. With that in mind, this section will discuss some of the limitations and potential areas for improvements.

6.2.1 Data limitations

As mentioned in section 3.2.2, the temperature data used in this analysis comes from a single sensor in the centre of Leicester. This may present a problem with some buildings if the local conditions around the sensor are different from the local conditions around the building in question. This problem could be eliminated by collecting temperature data at several points in Leicester and either merging the data or selecting the most appropriate (e.g. nearest) dataset for each building. This problem is expected to be minor as temperature usually varies only slightly within a city of similar elevations.

6.2.2 Core model limitations

The VBDD model used in this work is a simplification of reality. There are implicit assumptions of constant conditions such as free gains and infiltration rate. In practice these are unlikely to be met. In particular, the values of H_a, T_{in}, Q, η and f_{nh} used in equations 2.15 and 2.16 to describe the three VBDD model parameters will likely change continuously. For example, free gains, Q will be dependent on the number of people in the building and on the amount of solar radiation falling on the building.

With these variables changing over time the parameters estimated by OLS regression will represent the average values associated with the raw data and can only act as a guide. The effect can be minimised by ensuring the raw data is representative of the prevailing conditions (e.g. use a full year of daily values to cover all levels of solar gains).

The VBDD model is a very simplistic representation which is rarely applicable at resolutions higher than weekly resolution. This is mainly because buildings are occupied intermittently and control systems ensure that when a building is not occupied it is either not heated or is heated to a set-back temperature lower than the standard occupied temperature.

For example, if a building is closed during weekends then it may be maintained at a lower setback temperature. T_{in} will vary dramatically and predictably between the set point temperature and the setback temperature and when the model is fitted to real data the estimated parameters will be far from reality.

Applying the VBDD model with variable occupancy as described in section 4.1.2 goes some way to absorbing these problems. However, similar occupancy variations also occur on an annual basis. School buildings are the most obvious examples. Schools are occupied differently and have different consumption patterns during the school term and the school holidays.

Any attempt to analyse a school without taking holidays into account will result in events being detected for each holiday. This may be correct in that events are happening, however it may be preferable to include a new independent variable relating to the school term time. This would result in events only being detected if the term time or holiday time consumption patterns changed. This is left as a suggestion for further work.

Another potential problem is the confounding effects of thermal mass. An improvement in the method would be to use daily historic temperature described by Wright (Wright, Young et al. 2005) instead of average daily temperature. Daily historic temperature is determined recursively and includes a portion of the historic temperature from the previous day, which in turn includes a proportion of the previous day (and so on until the first datum). Wright suggests it may be possible to improve the fit considerably using this method.

6.2.3 Occupancy detection limitations

The consumption models are modified to take account of weekly variations in consumption. In particular these were introduced to account for the many cases where weekend consumption is lower than weekday consumption. Without taking the day of the week into account, the standard deviation of the model residuals is higher than it would otherwise be and this affects the sensitivity of event detection.

However, adding the full weekly [0123456] variation increases model complexity significantly. The number of model parameters is multiplied by seven and this multiplies the SBC penalty making the complex model only likely to be selected if it produces a much better fit to the data. There are many datasets where the full weekly model is not necessary, this is why the [0111110] and [0111112] models were also introduced.

These simpler models represent common consumption patterns but they by no means cover the whole range of potential patterns. Another pattern which may be common is one where Mondays or Fridays or both are different from other weekdays due to equipment being switched off during the week. Another example is Libraries in Leicester which are often closed on Thursdays.

Datasets exhibiting any pattern not specifically covered will either be modelled using the [0123456] variant and be given too many parameters or will be under-specified by a simpler model and suffer from a poor model fit as a result. To ensure the 'correct' model variant was used for all datasets using the method described in this work would mean fitting hundreds of alternative patterns and comparing them all with SBC. This is clearly not an ideal solution.

An alternative which might be attempted is to fit the [0123456] variant and compare the models for each day of the week with each other to see if the model parameters indicate they could be successfully merged. Using this kind of approach the modelling process could result in any of the potential patterns being fitted without an exhaustive search. Model optimisation may yield improvements in model fit but is outside of the scope of this work.

6.2.4 Event detection limitations

The event detection process itself has some idiosyncrasies which can lead to events being detected in error. The cause of this problem is the merging of the data before and after an event prior to the event being detected. Because the model fitting procedure is not aware of the event, certain events can ‘interfere’ with the event detection process.

To demonstrate this process, two years of temperature and consumption data covering two complete heating seasons were simulated with a single event in the summer half way between them. Temperature was modelled as a simple sine wave. Consumption was modelled using two VBDD models, one applied before the central summer period (sim 1) and the other applied after (sim 2). The models were identical except for an increase in the non-heating consumption and a decrease in the heating coefficient.

Figure 6.7 demonstrates the progression of event detection in these simulated data. The figure is split into four rows, each representing one step in the event detection process. On each row there are three charts. The first chart on the left shows the raw and modelled consumption plotted against temperature, this shows how the data are being modelled. The central chart shows the same data plotted against time, this is the best chart to show how effective the modelling is at picking out events. The chart on the right shows the CUSUM plot with 0.1% boundaries, this highlights the underlying event detection process.

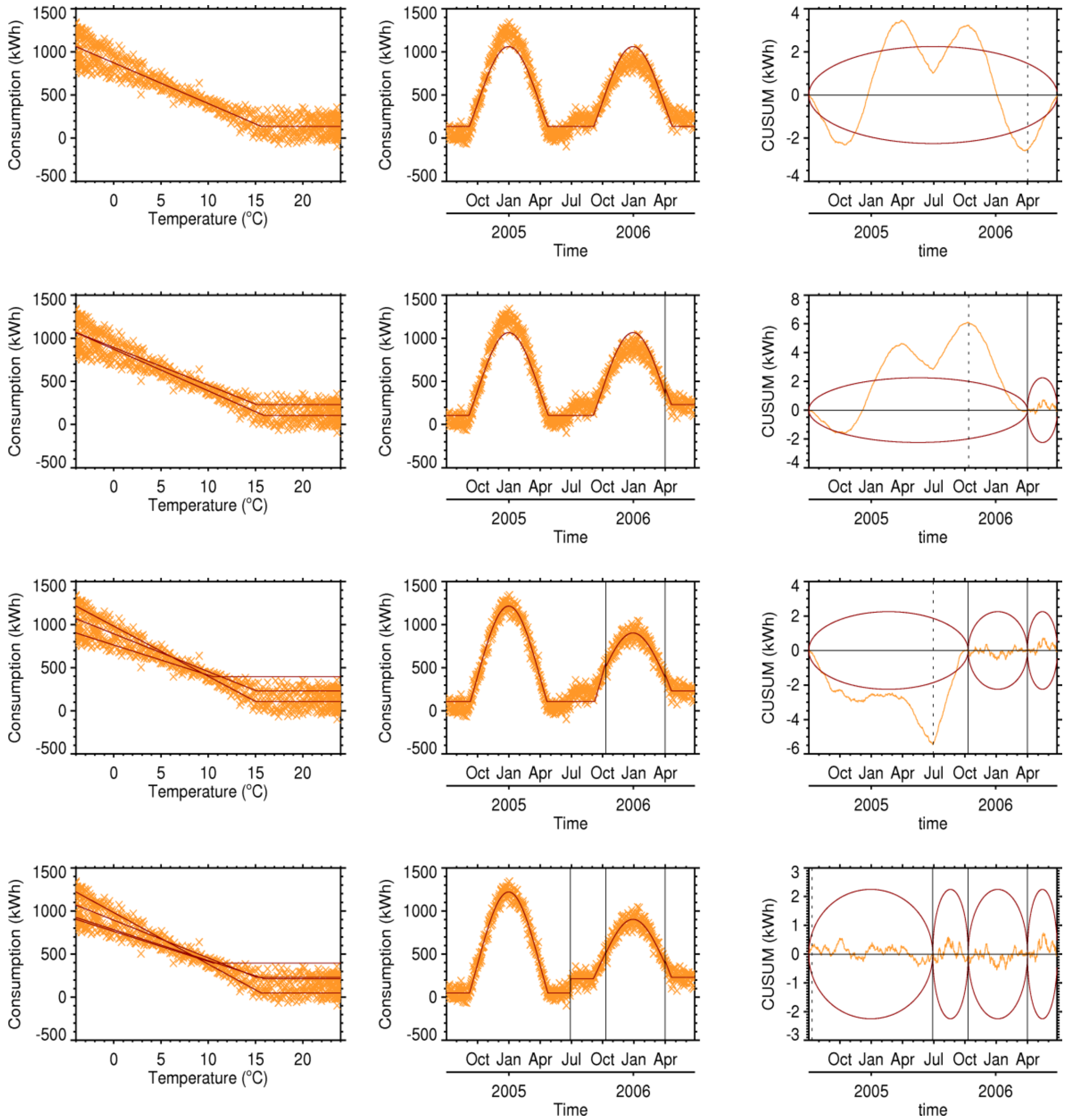


Figure 6.7: Interference in the event detection process

The first row shows the situation before any events are detected. In the left chart a single model is fitted through the whole dataset. The nature of the event is such that it causes the two models to 'cross over' on the temperature axis. That is, the simulated consumption for the two seasons cross over each other. The model generated by the dataset as a whole falls between the two simulated consumption patterns and cuts through both of them during moderate conditions in spring and autumn.

The effect of this can be seen in the middle chart. The model falls between the two patterns so it over-predicts in the first summer and under-predicts in the first winter.

The change occurs in the second summer and causes this pattern to flip such that the model over-predicts in the second winter and under predicts in the third summer.

This chronological switching between over-predicting and under predicting causes ‘phantom’ events to be detected where no real change has occurred. The CUSUM chart changes direction in October and April when the prediction crosses the consumption data. Even though the simulated event causes a sharp and strong signal in the CUSUM, it is outweighed by these ‘phantom’ events and an event is registered in the last April.

The effect is to remove a portion of the ‘sim 2’ data from the analysis. The second row shows the situation after this first ‘phantom’ event is added into the modelling process. A further ‘phantom’ event is detected in the October of the second year for the same reasons. This removes even more of the ‘sim 2’ data as show in the third row of the figure. At this point the simulated event is finally detected. As soon as the real event is detected the effects of crossing over are removed from the CUSUM chart. However, the ‘phantom’ events remain registered in the dataset and the final model is shown in the fourth row of the figure.

It seems that ‘phantom’ events are unavoidable with the OLS CUSUM approach to event detection. However, with simple post-processing they may be easy to remove. By definition these ‘phantom’ events are characterised by having no discernable influence on consumption patterns. One way to remove them might be to check each pair of neighbouring periods in turn to see if the event that separates them would be detected in isolation from the rest of the dataset. A true ‘phantom’ event would not be detected once the influence of the real event was removed from the analysis.

In the example this would happen after the final event was detected. We are left with three events to test. The first event would be tested by performing the event detection procedure on data from the beginning of the dataset up to the second event. The process would be repeated for the data from the first event to the third event and finally from the second event to the end of the dataset. Since these data are entirely simulated by two models, the event detection would only find the simulated event and would reject the ‘phantom’ events.

This approach can be applied to each dataset after the initial event detection process has identified ‘candidate’ events. Events will be either confirmed or rejected by the procedure. Events could also conceivably be moved by the procedure. If events are rejected or moved then the whole dataset should be tested again to ensure the effects of these changes are taken into account.

It may be the case that the spike in events that have only a very small impact on consumption levels shown in Figure 6.4

6.3 Opportunities and applications

The analysis generates a large amount of event-oriented meta-data consisting of individually modelled periods separated by events. Each dataset, analysis, event and period is recorded in a database and is available for inspection via SQL queries. These records combine to form a bottom-up description of changing consumption patterns across the entire building portfolio. There is huge potential for building tools for energy management information systems from this data. In this section some potential approaches are suggested.

6.3.1 Navigating seas of data

One problem that the raw consumption data presents is one of sheer volume. Several software packages exist which allow the user to view energy consumption data in charts and conduct simple analyses on individual datasets. But with dozens of datasets it becomes impractical to manually review consumption in every dataset individually on a frequent, regular basis. Event-oriented meta-data provides an elegant solution to this problem.

In the current work the meta-data are simply plotted in charts for interpretation as a whole. These data could alternatively form the basis for software to allow analysts to explore the (raw or cleaned) energy consumption dataset as a whole. This would be done using the meta-data as a guide. With an intuitive, interactive user interface, such a tool could provide an unprecedented level of insight into an entire portfolio of buildings.

Such software would include a number of different ‘views’ (or interactive reports) on the building portfolio which include functionality based on simple transformations of the event-oriented meta-data. Each view would be designed to give the user the information they need to perform a certain task. There are many possible views and they will not all be discussed here. For illustration of the main point, two simple views will be described.

A core feature is a view of a single, half-hourly dataset as a time series. This would be much like the standard energy consumption charts offered by existing energy management software but with a significant event-oriented enhancement. Each event would be 'bookmarked' in the data and links to each bookmark in the active dataset would be provided as part of the interface (e.g. in a drop-down menu).

The key functionality would be to plot the periods before and after the corresponding event when a link is selected. The event in question could also be highlighted and the user presented with data pertaining to the event and to the periods before and after the event. This allows the user to navigate a single dataset, skipping freely from event to event.

This navigation between events within one dataset would be very useful for understanding a given dataset and diagnosing problems. Several such views would be possible where the event and the periods of consumption either side of the event are the key subjects. One alternative is to present two charts of modelled consumption data (against time or temperature as appropriate), one for each period associated with the event.

Another valuable form of view that can be derived from the available data is one capable of navigation between events based on the event characteristics. Such a view could take the form of a league table of events, searchable and sortable by date, significance and impact. Simple filters could restrict the list to given subsets. The list could optionally be sorted by the magnitude of the worsened performance to provide an element of prioritisation.

Dynamic league tables would allow the user to browse through the most significant events, those with the largest impact and those in a given period of time. The list itself would be made of links to the appropriate bookmarks. Clicking an item in the list redirects to the dataset view described above with the appropriate event pre-selected.

A simple link to take the user back from the dataset view to the league table view would complete the simple application. With these two simple views the analyst is given the ability to navigate back and forth across the most important segments of a very large dataset representing a portfolio of buildings. All that is required is that the event detection analysis is conducted on a daily basis to keep the underlying meta-data up to date.

Of course this simple description does not describe a fully featured application. The example is given to illustrate the basic point that event-oriented meta-data provides a very handy means to focus the attention of analysts where it is most likely to yield results. In this case the term ‘results’ implies opportunities to investigate and diagnose problems, to develop potential energy efficiency interventions and to gather the information necessary to publicise successful interventions.

6.3.2 A portfolio-wide knowledge base

In a further development this approach lends itself to the addition of user generated information about each event. Adding structured (e.g. category, Boolean values) and unstructured (e.g. arbitrary notes) data would, for example, enhance the league table search functionality and provide a convenient store for information gathered during investigations into the causes of all types of events.

A good example of user generated data is a category value to represent the stage reached in event processing. This value could be set to ‘unprocessed’ by default and upgraded to ‘under investigation’ or ‘ignored’ once the analyst has viewed the event (depending on whether the event is considered a priority). Once an investigation has concluded the value could be changed to ‘diagnosed’ (or ‘ignore’ or ‘inexplicable’). If the diagnosis represents an opportunity to intervene then the value could be upgraded to ‘potential intervention’ and if corrective action is finally taken it could be set to ‘implemented’.

Events can be categorized in other ways too. Faults can be labelled as such, and linked to the equipment or technologies involved. The same is true for energy efficiency interventions. For example, if heating controls fail in a building then an event is detected and it is labelled as a fault and as being connected to heating controls.

Thus, if all events are systematically investigated and appropriately tagged then strategic reports indicating the impact of each technology could be generated. These would be simple lists of faults and interventions for each technology, including estimated impact and significance data. The total impact of events relating to each technology could be easily compared.

This data could be used to create a separate list of events for each category value. An analyst may be able to produce a list of ‘unprocessed’ events to browse. They can filter out data problems and known changes in use by assigning them to ‘ignore’. They can instigate investigations by assigning them to ‘under investigation’.

Similarly an engineer may be interested in those events in the ‘under investigation’ category and add notes against events as part of the ongoing investigation. In this way, each event gets specialist attention and more information is gathered. The event is pushed inexorably towards either the ‘ignore’ category (or similar) where no further action is required or into the ‘potential intervention’ category where it is added to the implementation plan.

It is also useful to make a distinction between events whose effects are permanent and those whose effects can be and are reversed over time. In terms of energy management, detecting a reversible deterioration in energy performance is highly desirable. In many cases, for example if heating controls are overridden, such changes can be reversed very simply. Tagging events as being reversed or linking two related events in this way could easily be done in a similar way.

Events whose effects are permanent are also of great interest from a strategic point of view. Understanding the kinds of changes which lead to permanent improvements and deteriorations allows for a clearer picture of the effects of strategic decision making. Permanent events, once diagnosed and understood could be tagged as such and the details collated into a strategic report.

Taking this idea even further it is possible, with careful thought, to develop event-oriented software which would closely follow the strategic approach described in section 1.3. Those events which present a potential intervention are tagged as such. Each opportunity generated in this way can be linked to the appropriate structured data necessary to make an assessment. Once an opportunity is implemented it can be updated to reflect this and even linked to the event which results from the intervention.

Chapter 7 Conclusions

"The winds and waves are always on the side of the ablest navigators"

– Edward Gibbon (1737 – 1794)

In Chapter 1 the research aims and objectives were set out in terms of developing a method for analysis of half hourly energy consumption data. This chapter completes the thesis by returning to these aims and objectives and reviewing to what extent they have been met. The chapter is set out in four sections. The first two sections deal explicitly with the aims and objectives: section 7.1 discusses the methodological background and how existing methods have been developed in this work; section 7.2 discusses the analysis of the case study dataset provided by Leicester City Council. Section 7.3 discusses the potential for further research. Section 7.4 provides a final summing up of the key message that can be taken from this work.

7.1 Method development

The primary aim of this work was to develop a method to model changing energy consumption patterns in buildings. In particular the method was necessary to take advantage of automatically generated high-resolution data. This aim has been achieved through a combination of traditional energy management analysis techniques, a novel consumption modelling process and statistical tests for structural change in modelled data.

The first objective of this research was to assess the state of the art in energy consumption data analysis. Chapter 2 provides a summary of the findings of this assessment. This research draws from existing techniques in the energy management literature which were uncovered in this process.

7.1.1 Monitoring and verification

Monitoring and verification (M&V) is a term which describes the processes and techniques used to verify savings attributable to energy efficiency retrofits. M&V techniques for comparing data from before and after an intervention are described in section 2.1.2. Savings calculations are, without exception, applied with respect to a known intervention. For example, a boiler retrofit might be carried out by an energy services company. Once sufficient time has passed an analysis would be conducted to estimate the savings due to the retrofit.

Techniques used in M&V such as the variable base degree day (VBDD) model (see section 2.2) and normalised annual consumption (NAC, see section 2.3) are important elements of this work. The main difference is that the savings calculation is applied to events of unknown origin which have been identified directly from analysis of the data themselves.

From an M&V perspective this may not have a significant impact on the status quo. It could be used to confirm that the intervention in question occurred at the reported point in time but this is unlikely to be a problem. It could also be used to disaggregate the effects of changes other than the intervention under study from the results.

Using the approach developed in this work for M&V would ensure the analysis was conducted from the basis of consistent patterns of consumption both before and after the intervention in question. If any unexpected changes in consumption occurred in the periods before and after the intervention then they would be identified and could be taken into account in the savings calculation.

7.1.2 Monitoring and targeting

Monitoring and targeting (M&T) is a term which describes the processes and techniques used to identify energy performance in buildings and to continuously monitor buildings for changes in that performance. Poor performance or deterioration in performance is indicative of waste. When changes in performance occur, they are investigated and corrective action is taken as necessary.

The standard CUSUM approach (see section 2.4) is used to identify when consumption diverges from expectation. Limitations of the technique are discussed in section 2.4.4. The present work overcomes these limitations. The monthly performance line is replaced with a variation of the VBDD model employed at daily resolution and CUSUM is automated using statistical tests for structural change.

The new approach is fully automated and represents a new way to conduct M&T analysis on a large scale. Not only does the analysis provide, in a matter of seconds, a far more detailed, accurate and repeatable version of what traditional techniques provide; it also provides something completely new in the event-oriented meta-data it generates.

Chapter 5 and Chapter 6 present a set of basic analyses that allow for individual datasets, periods within datasets or events to be picked out by virtue of their characteristics. Datasets of interest can be identified by their goodness-of-fit; model complexity; average period length; or overall performance improvement or deterioration. Periods of interest can be identified by their individual model parameters. Events of interest can be identified by their significance, their impact or their impact relative to the dataset in question or to the portfolio as a whole.

The opportunity here is clear; making use of these data within energy management will place the information, previously only held in high-resolution energy consumption data, at the finger-tips of energy managers enabling them to quantify historical events, prioritise investigations and easily assess the value of investments. All of these important tasks can be conducted automatically and in a few moments.

7.1.3 A novel modelling regime

A further objective of this work was to identify appropriate models to describe energy consumption patterns. To meet this objective a complex modelling regime was developed by combining existing models. The model is described in detail in section 4.1.

A defining characteristic of the modelling regime developed in this work is the value given to parsimony. Using the simplest model which captures consumption patterns in a given dataset avoids the generation of superfluous, meaningless model parameters and makes the model structure a more meaningful description of the consumption pattern. This makes the resultant meta-data more concise and more meaningful.

The final model includes over 128 potential model variants. Using a robust technique (Schwartz Bayesian Criterion, see section 4.1.4) to select between model variants has proved to be successful. Model fitting has revealed a lot about the patterns of consumption in the building portfolio.

The model type selected for a dataset is embedded in the meta-data such that it would be possible to identify, for example, all cases where an event caused the model to shift from a weekday/weekend model to one where the weekends were the same as the weekdays. Such events may be easily reversible control faults or behavioural changes.

7.1.4 Automated event detection

Another objective of this work was to determine a statistical method capable of identifying changes in energy consumption patterns. Consumption patterns are determined by fitting the consumption model as described above. Changes in this pattern are detected using OLS CUSUM (see section 2.5) to test for parameter instability. If an event is detected then the data are split into two periods. This testing and splitting is applied as a binary recursion (see section 4.2) to split the data into multiple periods of consistent consumption pattern.

The benefits of automated, objective event detection are significant. The statistical basis for the technique means the analysis is more reliable than the standard CUSUM analysis. Each detected event is known to have a minimum significance determined by the analyst and each event has a significance value associated with it. The analysis is entirely repeatable such that, for the same significance value, the same dataset will produce identical results time and time again, this was not the case with the standard CUSUM method where different analysts could produce widely differing results. In addition, the more sophisticated modelling regime provides detailed event-oriented meta-data describing the changing consumption patterns with each period between events being associated with a chosen model type and parameters.

7.2 Data analysis

Leicester City Council provided 738 half-hourly gas and electricity consumption datasets and outside air temperature data (see section 3.2). The data were collected from over 300 buildings in Leicester, UK. Event-oriented meta-data were generated from these data and stored in a database for further analysis. This section details the main findings from the analysis of these meta-data.

7.2.1 Consumption modelling

Another objective in this work was to prepare the data for analysis and apply the methodology to the 300 buildings for which data are available. The data were cleaned, interpolated them to daily values and analysed systematically.

The provided data were observed to contain many instances of erroneous and missing data points. Prior to analysis the data were cleaned of aberrant values (see 0). The data cleaning method was not entirely automated and is not strictly part of the methodology presented in this work. It was developed to handle problems with the particular dataset used in this work and may not be generally applicable. It is not known whether the data problems identified in the Leicester City Council data are common to all large scale metering systems or whether they are a symptom of the early adoption of untested technology. Either way, it is desirable to collect good quality data in the first instance rather than to rely on statistical means to clean data. That said, it is always important to check data for errors and the methods described in Appendix A provide a good basis for doing so.

The cleaned, interpolated datasets were fed into the analysis methodology to generate an event-oriented model of consumption for each individual dataset (see section 5.2). Meta-data were generated and stored in a database for each analysis.

An analysis of the goodness of fit (see Table 5.2) reveals that the core models provide a good fit to the data and that the parsimonious approach taken in this work is effective in keeping the consumption model as simple as possible without compromising on goodness of fit. The fully automated analysis including event detection provided a mean error of 60.9% and a median error of 31.6%. Well over half of the datasets are modelled to within the 35% tolerance set out in M&V publications for savings calculations.

A total of 5,269 events were detected separating 6,007 modelled periods. An inspection of the model types chosen for these periods by SBC reveals that the more complex models are rarely used and simpler patterns are far more common. Models with no weekly variation are the most common, followed by those where weekdays and weekends are modelled separately. These account for 76.5% of all individual periods of consumption modelled. Only 14.4% of periods were modelled with the most complex variant where all days of the week are treated independently.

The VBDD change point temperature was given as an example of inspecting model parameters across the modelled datasets. It was shown that, for the dataset under analysis, the most common change point was between 16ºC and 17ºC, significantly higher than the standard assumption for degree day weather normalisation and performance line calculations. This is thought to be due to either high internal temperatures (high thermostatic set-points), low internal gains or high heat-loss coefficients.

7.2.2 Demonstration

The final aim of this work was to demonstrate the value of this approach to energy management through detailed examples. Several individual examples have been described, it is the case that most examples are interesting and most show very clear changes, not only in the level of consumption but also in consumption patterns.

In the example in this work, over 45,000,000 half-hourly consumption records were reduced to just 6,007 individually modelled periods. Each period has up to twenty-one model parameters associated with it but typically no more than six. The model parameters describe the salient aspects of the consumption pattern in each period.

The analysis of aggregated meta-data provides a clear demonstration that the information extracted from energy consumption data in this work is valuable for both energy management purposes and for wider research into the building stock.

Beyond the energy manager, this approach to data analysis provides an opportunity to really get to grips with the nature of the changing building stock. To understand how often energy consumption patterns in buildings change, how much they change and what causes those changes to occur. It may never be possible to predict these changes but it may be possible to characterise and diagnose problems from the half-hourly patterns and changes in these half-hourly patterns.

7.3 Further research

An important use for event-oriented meta-data is in academic research conducted on the building stock. The analysis described in the present work provides an objective means to identify improvements and deteriorations in energy performance over time across large numbers of buildings.

Currently, the only way to understand changing buildings is to collect data on reported interventions and anecdotal evidence of changes in performance. Consumption data analysis may be conducted but usually with reference to reported information. It may be possible to use consumption data systematically and search for events manually in a small number of buildings but prior to this research no method existed to isolate changes automatically. The new approach allows for very large samples of buildings to be investigated.

The kind of information generated in the present work has significant benefits. In particular it is objectively determined from measured consumption data. If something causes consumption patterns to change then its effect will be isolated and quantified by this technique. This information was previously not available and is of great interest when studying energy consumption in buildings.

Given a large dataset, for example that which a utility company might maintain for billing purposes, it would be possible to automatically identify huge numbers of events across a wide variety of buildings. This would provide a strong foundation on which to build research projects to study the nature of change in the dynamic building stock.

By systematically investigating each event in turn it would be possible to identify, for example, how many events showing 'savings' were the result of intentional energy efficiency improvements and how many were the collateral effect of non-energy decisions. It would be interesting to identify what caused the decrease and whether the consumption has moved elsewhere, e.g. to another building or process.

This is just one example. There are many such studies which could use event-oriented meta-data as a starting point. The user-generated data added to each event and the post-processing would depend on the research question. It has the potential to form the basis of any number of rigorous studies of the changing building stock.

7.4 Final thoughts

It is now common for organisations such as local authorities to collect high-resolution energy consumption data from large numbers of buildings. Large databases of half-hourly resolution consumption data are already common and will most likely become the norm in years to come. The move from monthly billing data to half-hourly metering represents a huge increase in the volume of data to be managed and analysed. One year of monthly data includes only 12 values whereas one year of half-hourly data contains 17,520 values. The additional information available in these data provides a valuable resource for energy management. In particular, levels of weekend and overnight consumption often reveal areas of wastage hidden by monthly data.

However, the additional information comes at a cost. The process of manual inspection and analysis is cumbersome and, when dealing with several hundred half-hourly datasets, it is not practical to continuously review these data on a frequent basis. The extra effort involved can be an inefficient use of resources, particularly in cases when consumption patterns have not changed since the last inspection. In these cases, no new information is gained from looking at the data yet again. Manually inspecting data in this way inevitably leads to an inefficient use of resources.

It is not expected that consumption patterns will remain constant in the long term. Changes to consumption patterns represent changes to the underlying energy systems and must be investigated and managed. Events which cause such changes are of critical interest to energy management. The techniques described in this work enable these data to be automatically processed and split into periods of consistent consumption. Wherever a change in consumption pattern occurs, it is identified and the data are split. This removes the need for regular manual inspection as the analyst can simply browse the records of events to isolate recent or important events based on their characteristics. The automated method also identifies changes in cases where events cause subtle changes to consumption patterns. These may easily be missed by simple visual analyses.

The process is entirely automated and repeatable. Raw data are processed and converted into event-oriented meta-data. It is feasible for any local authority or similar multi-site organisation with the appropriate data collection systems in place to conduct this analysis and generate event-oriented meta-data for their own buildings.

The method is not perfect. There are limitations in the core models used which provide a greatly simplified view of consumption patterns. This is partly a limitation in the available data. If appropriate data were available, this problem could be reduced through the addition of alternative models.

The method benefits from being extensible. The modelling regime can easily absorb more model types since it includes a method for choosing the ‘best’ model for each dataset. This is currently achieved via a comparison of the SBC of each model. Adding additional model types would make the analysis capable of accurately describing more datasets but would also add to the computational burden.

Including a model selection step in the analysis means that where consumption patterns are simple, the model parameters reflect that and conversely, where more sophistication is needed to describe consumption, the method will produce a more complex result. The analysis reveals the weekly occupancy pattern for each period of consistent consumption. It also tells the analyst whether there is any variation with weather and produces a normalised estimate of annual consumption.

However, the model selection process is not perfect as not all the possible models are compared. This is a compromise between getting the absolute 'best' model and testing every model in the available set. Testing all models is computationally expensive at a large scale but a more sophisticated model selection algorithm could produce a better result.

The benefits of working with event-oriented meta-data in conjunction with the original half-hourly data are numerous and significant. The data provide an invaluable insight into the historical consumption patterns of a building at a glance. For a given dataset the model can be inspected easily by simply plotting the prediction against measured data.

The analysis works with the assumption that consumption patterns change at distinct points in time. Though this may be the case sometimes it is unlikely to always hold. There will be many cases, such as major retrofit work, where interventions are implemented over several days. In these cases the event detection methodology will generate several events over the period of change.

In addition, there are issues of 'phantom' events being generated in cases where consumption patterns change in certain ways. In particular this can happen where models 'cross over' each other on the temperature axis. Events are generated where no change has occurred. These can be addressed by further processing the meta-data to isolate and remove any 'phantom' events.

It is when many datasets are analysed together that a more powerful interrogation is possible. Events and periods can be compared across datasets to identify the relatively small number of points in the vast dataset which present the most useful information for energy management. Crucially, the meta-data are structured and meaningful and so can be interrogated to reveal changing consumption patterns across an entire portfolio. They can be inspected to extract general information about the portfolio as a whole or to look at specific datasets or subsets of datasets. They can be used to isolate datasets, periods and events of particular interest.

For energy management these data are useful in that they provide information about historical events across the portfolio. This information may be crucial to understanding the history of a given building, to place the current performance in the proper historical context and in managing future interventions. The regular analysis of data in this way can warn of unexpected faults, aid in the commissioning of interventions and quantify the effect of unexpected, non-energy (collateral) decisions.

References

Local Government Act 1999, HMSO.

Abrahamse, W., L. Steg, et al. (2005). "A review of intervention studies aimed at household energy conservation." Journal of Environmental Psychology **25**: 273-291.

ASHRAE (2005). Energy estimating and modeling methods. ASHRAE handbook - Fundamentals, American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Bennet, C. J. and M. A. Whiting (2005). Navigating Energy Management: A Roadmap for Business, The conference board.

BRECSU (1993). GPG 167: Organisational aspects of energy management, Building Research Energy Conservation Unit.

BRECSU (1996). GIL 50: Energy saving in buildings - methods for quickly identifying opportunities, Building Research Energy Conservation Unit.

BRECSU (1996). GPG125: Monitoring and Targeting in small and medium-sized companies, Building Research Energy Conservation Unit.

BRECSU (1998). GPG112: Monitoring and Targeting in large companies, Building Research Energy Conservation Unit.

BRECSU (1998). GPG125: Monitoring and Targeting in small and medium-sized companies, Building Research Energy Conservation Unit.

BRECSU (1998). GPG231: Introducing information systems for energy management, Building Research Energy Conservation Unit.

CIBSE (2006). Technical Memorandum 41: Degree-days: theory and application, The Chartered Institution of Building Services Engineers.

CIBSE (2009). Technical Memorandum 39: Building energy metering, The Chartered Institution of Building Services Engineers.

Claridge, D. E. (1998). "A Perspective on Methods for Analysis of Measured Energy Data from Commercial Buildings." Journal of Solar Energy Engineering **120**(3): 150-155.

Day, A. R., I. Knight, et al. (2003). "Improved methods for evaluating base temperature for use in building energy performance lines." Building Serv. Eng. Res. Technol. **24**(4): 221 - 228.

Dhar, A., T. A. Reddy, et al. (1999). "A fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable." Journal of Solar Energy Engineering **121**: 47 - 53.

Efficiency Valuation Organisation (2007). International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings, Volume 1.

Enders, W. (2004). Applied Econometric Time Series, Wiley.

Energy Metering Technology Ltd. (2010). "Dynamat data analysis system." from http://energymeteringtechnology.com/products/dynamat/.

Energy Savings Trust. (2011). "The Nottingham declaration on climate change." Retrieved 18/02/2011, 2011, from http://www.energysavingtrust.org.uk/nottingham.

EU (2006). DIRECTIVE 2006/32/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 5 April 2006 on energy end-use efficiency and energy services and repealing Council Directive 93/76/EEC, European Union.

European Commission. (2011). "Covenent of mayors." Retrieved 18/02/2011, 2011, from http://www.eumayors.eu/home_en.htm.

Farouz, S., J. C. Baltazar-Cervantes, et al. (2001). Monitoring and verification procedures used in the Texas loanstar and rebuild America programs, Energy Systems Laboratory, Texas A&M University.

Fels, M. (1986). "PRISM: An introduction." Energy and Buildings **9**: 5 - 18.

Firth, S., K. Lomas, et al. (2008). "Identifying trends in the use of domestic appliances from household electricity consumption measurements." Energy and Buildings **40**: 926-936.

Grubbs, F. (1969). "Procedures for Detecting Outlying Observations in Samples." Technometrics **11**(1): 1 - 21.

Haberl, J. S., C. Culp, et al. (2002). ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine What Was Really Saved by the Retrofit, Energy Systems Laboratory, Texas A&M University.

Haberl, J. S., A. Sreshthaputra, et al. (2003). "Inverse Modeling Toolkit: Application and testing." ASHRAE transactions **109**(2).

Haberl, J. S., W. D. Turner, et al. (2002). LoanStar after 11 years: A report on the successes and lessons learned from the loanStar program, Energy Systems Laboratory, Texas A&M University.

Harris, P. (1989). Energy monitoring and target setting using CUSUM. Cambridge, Cheriton Technologies Publication.

Hernandez, P., K. Burke, et al. (2008). "Development of energy performance benchmarks and building energy ratings for non-domestic buildings: An example for Irish primary schools." Energy and Buildings **40**: 249-254.

Hooke, J. H., B. J. Landry, et al. (2003). Energy Management Information Systems ACHIEVING IMPROVED ENERGY EFFICIENCY A handbook for managers, engineers and operational staff, The Office of Energy Efficiency of Natural Resources Canada.

ICLEI (2006). Cities for Climate Protection Campaign - The Reinforced Strategy for Europe, International Council for Local Environmental Initiatives.

IPCC (1990). Impacts Assessment of Climate Change – Report of Working Group II, Intergovernmental Panel on Climate Change.

IPCC (1990). The IPCC Response Strategies – Report of Working Group III, Intergovernmental Panel on Climate Change.

IPCC (1990). Scientific Assessment of Climate change – Report of Working Group I, Intergovernmental Panel on Climate Change.

Karatzas, I. and S. Shreve (1991). Brownian Motion and Stochastic Calculus. New York, Springer.

Katipamula, S., T. A. Reddy, et al. (1998). "Multivariate regression modeling." Journal of Solar Energy Engineering **120**: 177 - 184.

Kissock, J. K. and C. Eger (2008). "Measuring industrial energy savings." Applied Energy **85**(5): 347 - 361.

Kissock, J. K., J. S. Haberl, et al. (2003). "Inverse Modeling Toolkit: Numerical Algorithms." ASHRAE transactions **109**(2).

Kissock, J. K., T. A. Reddy, et al. (1998). "Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings." Journal of Solar Energy Engineering **120**: 168 - 176.

Kreider, J. F., D. E. Claridge, et al. (1995). "Building energy use prediction and system identification using recurrent neural networks." Journal of Solar Energy Engineering **120**: 161 - 166.

Kreider, J. F., D. Cohen, et al. (1998). "Estimation of energy savings for building retrofit using neural networks." Journal of Solar Energy Engineering **120**: 211 - 216.

Lee, A. H. W. (2000). "Verification of electrical energy savings for lighting retrofits." Energy conversion and management **41**: 1999-2008.

Leicester City Council. (2011). "Leicester City Council - Energy Management." Retrieved 12/01/2011, 2011, from http://www.leicester.gov.uk/your-council-services/ep/energy-efficiency/energy/leicester-city-council-energy-management/.

Levermore, G., J (1992). Monitoring and targeting. Building energy management systems: An application to heating and control, E & FN SPON.

Mahdavi, A., K. Orehounig, et al. (2007). ANALYZING TRADITIONAL BUILDINGS VIA EMPIRICALLY CALIBRATED BUILDING PERFORMANCE MODELS. Building simulation 2007.

Page, E. S. (1954). "Continuous inspection schemes." Biometrika **41**: 100 - 115.

Page, E. S. (1961). "Cumulative sum charts." Technometrics **3**(1): 1 - 9.

Ploberger, W. and W. Krämer (1992). "The CUSUM test with OLS residuals." Econometrica **60**(2): 271-285.

Ruch, D., L. Chen, et al. (1993). "A change-point principal component analysis (CP/PCA) method for predicting energy usage in commercial buildings: The PCA model." Journal of Solar Energy Engineering **115**: 77 - 84.

Schwarz, G. (1978). "Estimating the dimension of a model." The Annals of Statistics **6**(2): 461 - 464.

Shewhart, W. A. (1938). "Application of statistical methods to manufacturing problems." Journal of the Franklin Institute **226**(2): 163-186.

Stuart, G., P. Fleming, et al. (2007). "Rapid analysis of time series data to identify changes in electricity consumption patterns in UK secondary schools." Building and Environment **42**: 1568-1580.

The Carbon Trust (2005). CTC506: Carbon Management, assessing and managing business responses to climate change, The Carbon Trust.

The Carbon Trust (2005). GPG352: An introductory guide to energy performance assessment, The Carbon Trust.

The Carbon Trust (2006). CTG004: Degree days for energy management — a practical introduction, The Carbon Trust.

The Carbon Trust (2006). CTX601: Introducing Local Authority Carbon Management, The Carbon Trust.

The Carbon Trust (2006). CTX602: Introducing Higher Education Carbon Management, The Carbon Trust.

The Carbon Trust (2006). GPG310: Degree days for energy management - a practical introduction, The Carbon Trust.

The Carbon Trust (2007). CTC713: Advanced metering for SMEs: Carbon and cost savings, The Carbon Trust.

The Carbon Trust (2007). CTV022: Energy management strategy, The Carbon Trust.

The Carbon Trust (2008). CTG008: Monitoring and targeting techniques to help organisations control and manage their energy use, The Carbon Trust.

The Carbon Trust (2008). CTX606: Introducing NHS Carbon Management, The Carbon Trust.

United Nations (1992). United Nations Framework Convention on Climate Change, United Nations.

United Nations (1998). Kyoto protocol to the United Nations Framework Convention on Climate Change, United Nations.

USDOE (2000). Measurement and Verification Guidelines for Energy Savings Performance Contracting, Federal Energy Management Program.

USDOE. (2004). "Guidelines for Energy Management." Retrieved 9/11/09, 2009, from http://www.energystar.gov/index.cfm?c=guidelines.guidelines_index.

USDOE (2008). M&V Guidelines: Measurement and Verification for Federal Energy Projects Version 3.0, Federal Energy Management Program.

USDOE. (2011). "Energy Star Benchmarking starter kit." from http://www.energystar.gov/index.cfm?c=evaluate_performance.bus_portfolio manager_benchmarking.

USDOE. (2011). "Federal Energy Management Program." Retrieved 18/02/2011, from http://www1.eere.energy.gov/femp/.

van Wijk, J. J. and E. R. van Selow (1999). Cluster and Calendar based Visualisation of Time Series Data. IEEE Symposium on Information Visualisation (INFOVIS'99). San Francisco.

Vesma, V. (2011). "Forthcoming events on energy management." from http://www.vesma.com/events.htm.

Vesma, V. (2011). "UK free dagree days." Retrieved 27/05/2010, from http://www.vesma.com/ddd/.

Warren, A. (2003). The Energy End-Use Efficiency and Energy Services directive: A summary of its objectives and contents, Association for the conservation of energy.

Wright, A., A. Young, et al. (2005). "Dwelling temperatures and comfort during the August 2003 heat wave." BUILDING SERV ENG RES TECHNOL **26**(4): 285-300.

Zeileis, A. (2003). Testing for structural change: theory, implementation and applications. Dem Fachbereich Statistik. Dortmund, Universitat Dortmund. **PhD**.

Zeileis, A. (2004). "Alternative Boundaries for CUSUM Tests." Statistical Papers **45**(1): 123 - 131.

Zeileis, A. (2008). Structural change tests for energy management. G. Stuart.

Zeileis, A., C. Kleiber, et al. (2003). "Testing and dating of structural changes in practice." Computational Statistics & Data Analysis **44**: 109-123.

Appendix A Data preparation

"If you want to be a good saddler, saddle the worst horse; for if you can tame one, you can tame all."

– Socrates (c. 469 –c. 399 BCE)

The raw data used in this work have been provided as simple meter readings. Chapter 3 describes this basic data format and describes the basic processing necessary to convert meter readings into useable consumption figures.

However, in many cases the data have been identified as problematic. This appendix identifies and describes two main problems: missing data (section A.1) and erroneous data (section A.2).

The appendix continues with a description of the processes undertaken to assess data for quality and the data cleaning processes developed to handle these problems.

It should be noted that, although these problems occur in fairly large numbers, with such a vast dataset the rate of occurrence of data problems is very low. The result of the processing described in this appendix is standardised data ready for further analysis.

A.1 Missing data

Figure 3.5 in chapter Chapter 3 shows the raw meter reading data for the electricity meter in a school kitchen. On first inspection the data seem to be in order and no significant problems can be identified. However, a closer look at this dataset reveals a problem that is common throughout the dataset under analysis.

Figure 7.1 shows the calculated number of pulses between each pair of meter readings plotted against time. It might initially seem as though this is an accurate reflection of consumption patterns since detail such as school holidays are now clearly resolved. However, the time step between readings is not always 30 minutes.

Variations in the time step need to be taken into account when looking at the energy consumption between readings. The number of pulses which occur between a pair of readings is dependent, amongst other things, on the length of time between the readings.

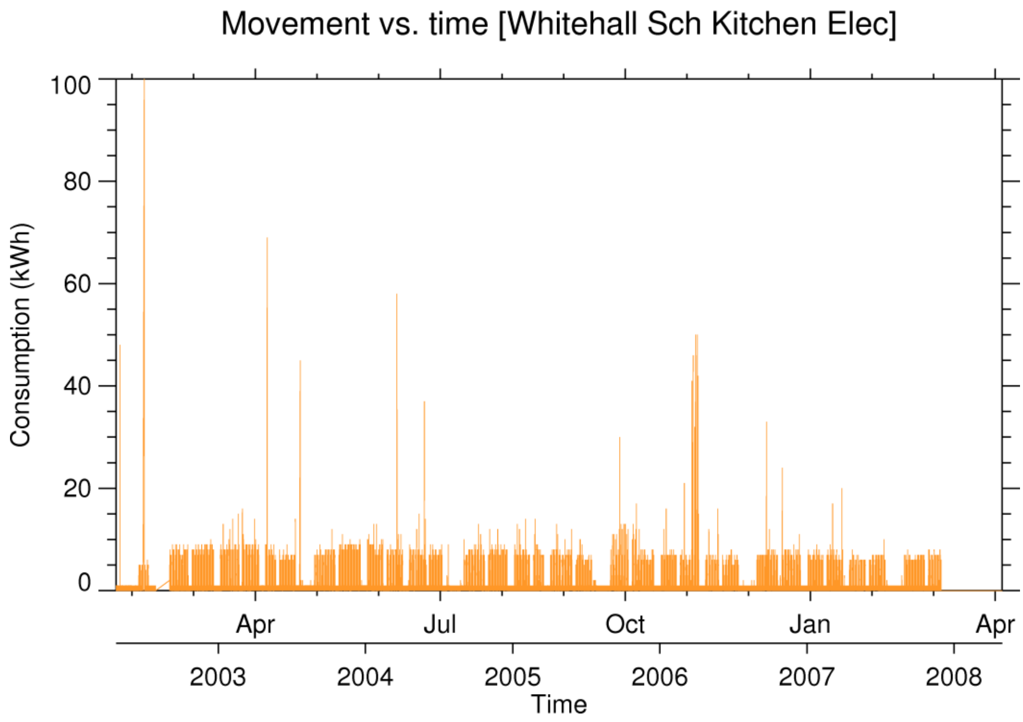


Figure 7.1: Example energy consumption data

In Figure 7.1 there are several periods where consumption appears to increase to several times the 'normal' maximum. This is not caused by high consumption during those times or by data errors. It is the effect of missing readings.

When a reading is missing in regularly spaced data it leaves a gap twice as wide as the normal gap between readings. So twice as much time has passed and the consumption value recorded for that large gap represents consumption which has taken place over twice the normal period. Thus, a gap in the data can be expected to be followed by a spike in consumption.

So a variable time step means that the data must be carefully processed before analysis. To ensure a fair comparison of consumption patterns over time and between buildings. There are two approaches utilised in the following sections of this chapter they are briefly introduced below.

One approach is to calculate the gradient of the cumulative pulse count (i.e. the average rate of consumption or average power). This produces a dataset which maintains the variable time step but normalises consumption over time. This is used in section A.4 to identify outliers in the meter reading data.

An alternative approach is to interpolate meter readings to a fixed time step. This produces nicely formatted data (i.e. with a common time step) but the actual meter reading values are lost and replaced with values based on assumptions about the pattern of consumption. For example, linear interpolation will assume the rate of consumption between two readings is constant. This is used in section A.3 to generate standardised datasets.

A.2 Erroneous data

It has been observed through simple visualisation that datasets occasionally include erroneous data. That is, data which are not valid meter readings but the product of some form of data corruption.

These problem readings will confound any attempts at analysing consumption patterns and so must be identified and removed prior to conducting any deeper analysis.

Errors have been observed in a large number of datasets, a thorough investigation into the extent of these problems is presented in section A.4. This section presents an introduction to the types of errors that have been observed.

Two general classes of error have been identified. In their raw form these are referred to as spikes and steps. Spikes are the result of individual erroneous readings where subsequent readings are unaffected. Steps involve a change which affects all subsequent readings. Figure 7.2 shows a small sample of datasets displaying both types of error and is to be compared with an unaffected dataset in Figure 3.5.

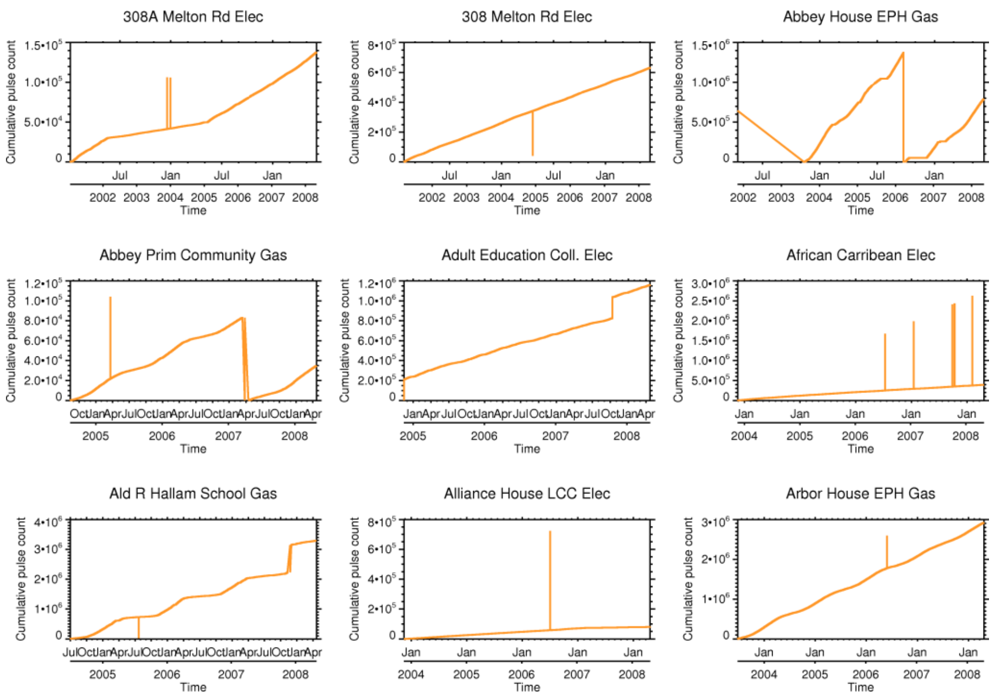


Figure 7.2: Meter readings for datasets with erroneous data

Spikes are so named because their most common form causes a spike in the raw metering data. However this is not always the case. Spikes are defined in this work as being the result of a single erroneous meter reading record. Their effects can be minimised by simply identifying and removing the offending record thus rendering the problem equivalent to a simple missing reading.

Spikes can be caused by corruption of the pulse count, timestamp or meter reference. As such they can lead to the wrong pulse count in the right place or the right pulse count in the wrong place.

Most spikes appear as individual meter readings which are extreme compared to real values at that point in the dataset. This common form of spike can be positive or negative and is apparent in all of the examples in Figure 7.2. Some spikes will also be associated with a missing reading either elsewhere in the same dataset or in a completely different dataset.

A less common form of spike appears as a single point either before the real data begin or after the real data have ended. The third (top-right) example in Figure 7.2 demonstrates and example of this where the first reading occurs over a year before the second and is clearly the wrong value. These must be the result of either a timestamp or meter reference error.

Spikes which have a value above the maximum pulse count for the dataset in which they appear cannot be the result of a timestamp error but may be caused by a corrupted meter reference or pulse count. This can be seen in the fourth, sixth and eighth examples (middle-left, middle-right and bottom-centre) in Figure 7.2.

Occasionally it seems that a meter reference error causes data to be repeatedly transferred between two datasets. This can be identified as a series of spikes in one dataset which appear to come from another dataset (i.e. they form a series of steadily incrementing values). This can be seen in the sixth (middle-right) and possibly in the first (top-left) examples in Figure 7.2.

Steps are more complex than spikes. In a step the pulse count jumps inexplicably and does not return back to the previous level as it does in a spike. The effect is similar to, and in some cases could be related to the effect of a meter 'rolling over' from the maximum possible value to zero (e.g. 999999 to 000000) as can be seen in the third and fourth examples (top-right and middle-left) in Figure 7.2. However, as shown in Figure 7.2, steps can be both positive and negative so ‘rollover’ cannot account for all instances.

Simply removing the meter reading associated with a step will not solve the problem as it does with a spike. There is a distinct shift in the value of the meter reading so removing a single reading will move the extreme increment over to the next reading. However, steps can be removed from meter readings by taking the derivative, removing the offending value and taking the integral thus generating a new (cleaned) set of meter readings.

A brief discussion of the potential causes for the observed data problems is presented in section A.5.1. Whatever their cause, these data problems cannot be allowed to pass through into analysis as they often dominate a dataset representing consumption levels orders of magnitude greater than the expected value. Removing the problems is not difficult but isolating erroneous data from high consumption is more problematic. A basic methodology to standardise data and remove errors is described in sections A.3 and A.4.

A.3 Standardisation

It has been shown that, although ostensibly recorded at half hourly intervals, meter readings in the dataset under analysis are better described as being ‘mostly’ recorded at ‘nearly’ half hourly intervals. This section provides a brief description of the extent to which this is a problem and how it has been addressed in this work.

There is a requirement (as discussed in chapter Chapter 3) to match metered consumption data with measured outside air temperature data. Since the timing of meter readings is not consistent across datasets it is necessary to interpolate datasets onto a regular half hourly time series. With a common time step across all datasets, direct comparison is possible.

Treatments such as linear interpolation provide a convenient means to place data on a common timescale with a fixed time step. The process inevitably affects the raw data, especially when gaps between readings are large. However, total consumption is not affected and for the most part the benefits outweigh the costs of the operation.

A.3.1 Variable time interval

As described in section 3.2.1 the actual point in time when meter readings are taken is not consistent across datasets. This is an intentional design feature of the system since it spreads the network traffic across each half hour.

Section A.1 shows that occasionally meter readings are missing leading to time intervals greater than 30 minutes. The cleaning process described in section A.4 also results in individual readings being removed which in many cases will leave gaps.

In addition, a simple scan of any dataset will reveal that the time between 'half hourly' readings is not always precisely 30 minutes. It can vary by up to 15 seconds either way (i.e. between 29 minutes and 45 seconds and 30 minutes and 15 seconds). The reason for this is not clear.

All these inconsistencies mean that data must be interpolated in order to establish both how much energy was consumed and what the average temperature was during a particular time period.

A.3.2 Data density

The average readings per day, referred to henceforth as the data density or ρdata is defined as the number of meter readings in a dataset, n divided by the range between the date of the first reading (t0) and last reading (tn-1) in that dataset.

$$\rho_{data} = \frac{n}{(t_{n-1} - t_0)}$$

The unit of data density is meter readings per day (written as day-1). For half hourly data the data density is expected to be 48 day-1 or a little less if a few readings are missing. Data density is a simple means to gain an overview of the extent of missing data.

The data density was calculated for each dataset under analysis. The distribution of data density between 0 and 60 day-1 can be seen in Figure 7.3. It is clear that most datasets exhibit a data density close to (within about 5%) the expected value and that there are significant numbers of datasets with between 5% and 20% missing data.

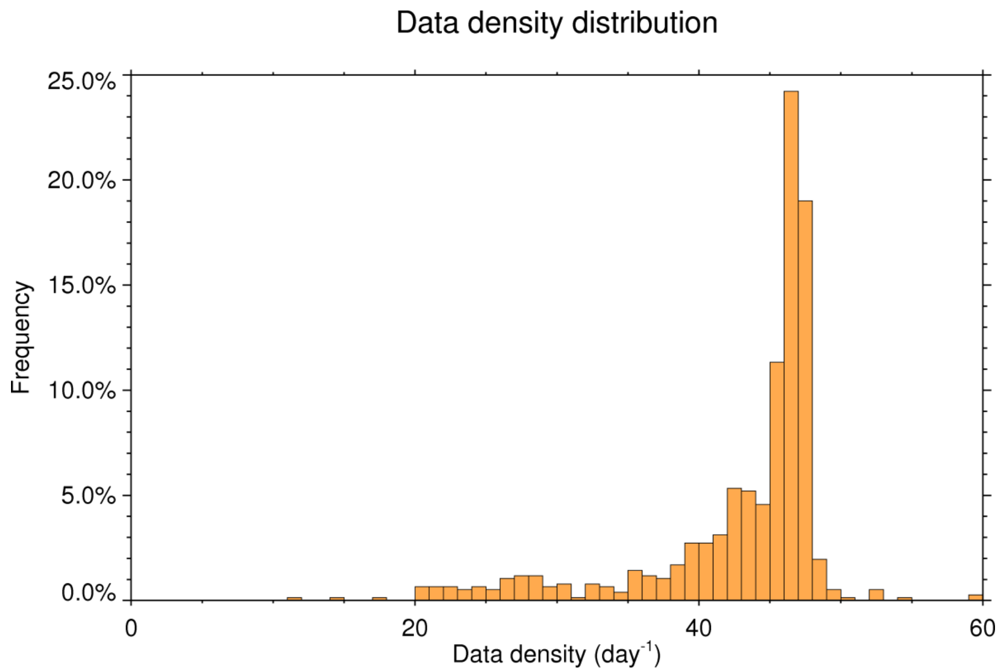


Figure 7.3: Data density distribution

There are 36 datasets with a data density of greater than 48 day-1. Sixteen of these are over 60 day-1 and are not shown on the chart. The maximum data density recorded was 133 day-1. On inspection these more dense datasets were found to have been recording meter readings more frequently for some periods and so do not pose a significant problem for analysis.

Several datasets exhibit a very low data density. The primary cause is contiguous periods where no metering data are recorded. This could be due to a failure of the communications network or of the meter itself. These reduce the number of meter readings within the period covered thus reducing the density figure. These instances are problematic and are handled as described in section A.3.3.

Low data density can also be caused by readings with erroneous timestamps as described in section A.2. Such data problems can artificially increase the range in the data density calculation and lead to a very low density figure. To combat this problem a simple filter for large gaps at either end of the dataset is applied.

A simple algorithm is applied before the cleaning processes described in section A.3. If the first reading is more than a day before the subsequent reading then it is removed. This process is applied iteratively until the first two readings are separated by no more than a day. This is repeated for the last two readings in the dataset.

A.3.3 Interpolation

Linear interpolation can be used to estimate what a meter would have recorded had it been read at any given time. Thus interpolation provides a convenient means to produce ‘standard’ half hourly datasets with readings every half hour, on the half hour.

Standard datasets have several benefits, primarily they are easy to compare since each reading is at the same point in time. This makes it easier to link consumption to other data such as outside air temperature.

To estimate a new pulse count, p (p0<p<p1) at time, t (t0<t<t1) between two meter readings (t0, p0) and (t1, p1) the following simple formula is used.

$$p = p_0 + (t - t_0)\frac{p_1 - p_0}{t_1 - t_0}$$

Linear interpolation in this case relies on the assumption that the rate of consumption is constant between meter readings. This is clearly not the case but is acceptable when interpolating one or two values over short time periods.

For half hourly data with missing periods where any gaps between readings are within a few hours linear interpolation provides an effective mechanism for infilling these missing periods. Figure 7.4 shows consumption values interpolated to half hourly periods for dataset presented in Figure 7.1. Missing data identified in section A.1 have been smoothed out by the process and all data points now appear exactly on the half hour.

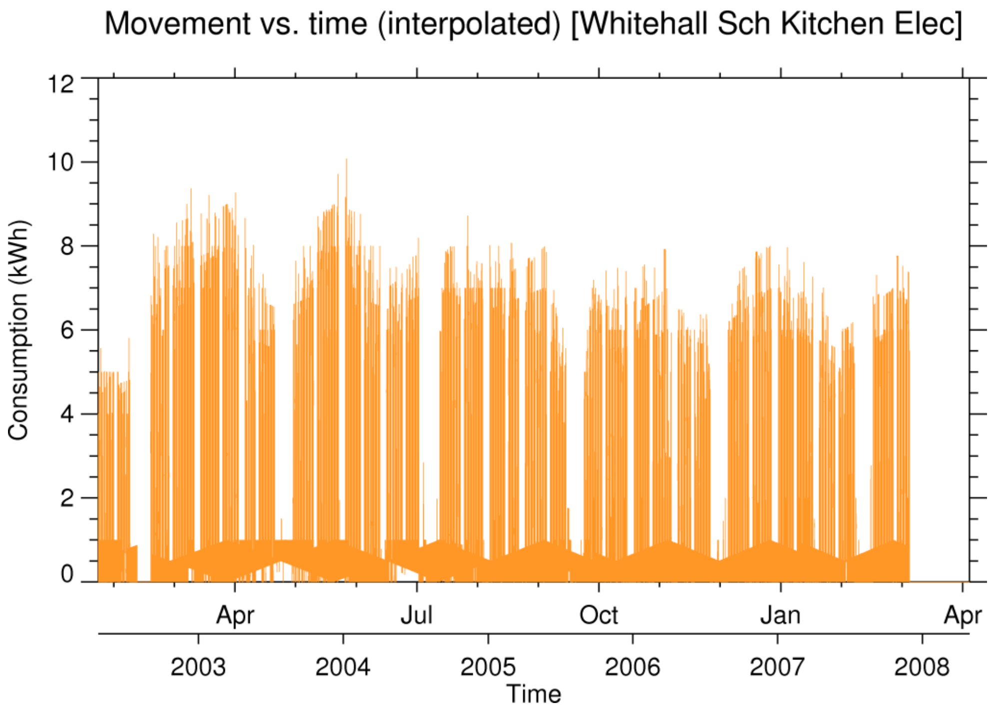


Figure 7.4: Typical interpolated consumption data

In some datasets, gaps between readings can be much larger than a few hours. This problem cannot be solved by interpolation alone. In fact, since the data are missing, there is no information about the detailed pattern during these periods.

Interpolation becomes less acceptable when the gap between readings is longer. Especially when the expectation is that the rate of consumption will vary over the period (day/night, weekday/weekend and seasonal patterns are all common). In this analysis, interpolation is used to estimate meter readings at half hourly intervals and at daily intervals.

In this work, large gaps are interpolated and flagged as missing data. Missing data do not contribute to the modelling process. This can introduce an element of bias but since the data are missing, they cannot be reconstructed in any way that would support the analysis.

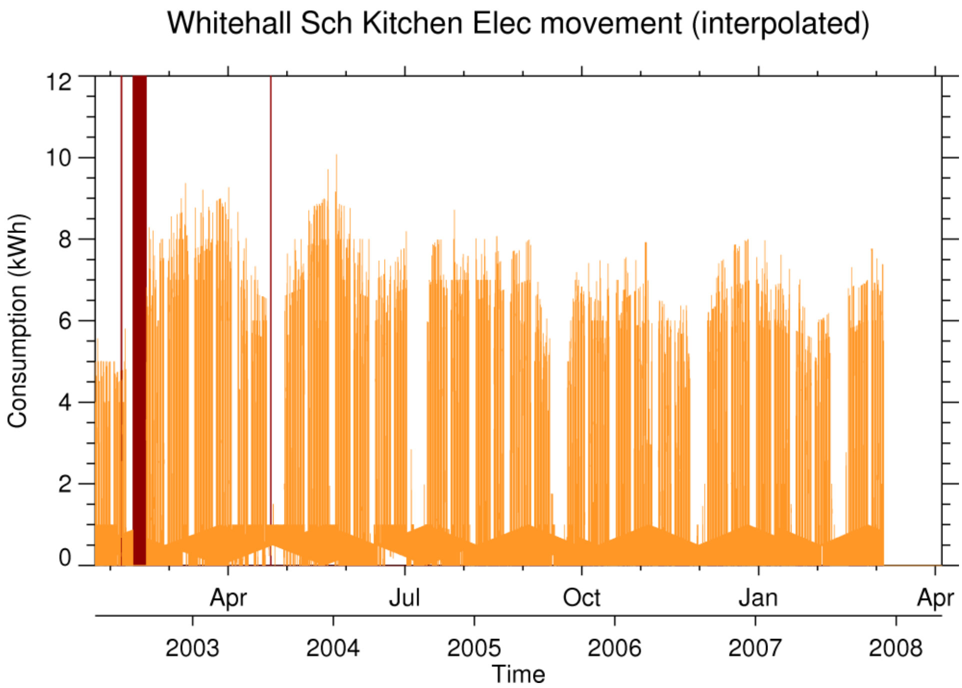


Figure 7.5: Typical interpolated data with missing data highlighted

A threshold has been set such that gaps between actual readings of 24 hours or less are ignored and interpolated 'silently'. Gaps over the 24 hour threshold are interpolated but the interpolated readings are flagged as 'missing'. Figure 7.5 shows the example dataset with missing data highlighted. In this case there were three gaps greater than 24 hours, the largest being about a month between readings in the summer of 2002.

A.4 Data cleaning

It is not difficult to remove erroneous data once they are identified. Section A.2 describes methods for doing so for both spikes and steps. This section describes techniques employed in this research to identify errors in metering data.

A robust method for identifying outliers needs to distinguish between erroneous data and genuine periods of high consumption. Most importantly, removing genuine data is to be avoided. Periods of highest consumption are most vulnerable to being falsely identified as erroneous. They are also of most consequence in terms of cost and emissions and may be indicative of wastage.

Distinguishing between erroneous readings and excessive consumption is particularly difficult with complex data such as whole building energy consumption. Common events such as using a kettle may lead to sporadic, isolated spikes which may share some characteristics with erroneous data.

A.4.1 Average rate of consumption

The approach taken in this work looks for extreme outliers in the average rate of consumption, r measured between each consecutive pair of meter readings. The rate of consumption or average power (measured in kW) is equivalent to the gradient of the meter reading graph. The change in pulse count, (p_i – p_{i-1}) is adjusted for meter calibration.

$$r_i = mk\frac{(p_i - p_{i-1})}{(t_i - t_{i-1})} \qquad (\mathrm{i} > 1)$$

where m is a multiplier representing the calibration of the meter (e.g., if the meter measures 0.1 kWh per pulse, m = 0.1), k is the coefficient used to convert between the measured units and target units (i.e. to convert kWh to kWh, k = 1), p_i and t_i are the pulse count and timestamp for the ith meter reading and ri is the rate of consumption for the period between t_{i-1} and t_i.

By calculating the rate of consumption over time, any variation associated with the variable time step is absorbed. Interpolation is an alternative method for normalising the time step but is inappropriate at this stage in the analysis. For error detection the average power is preferable as it maintains the direct link with the original meter readings. Once identified, the effects of erroneous readings can be removed.

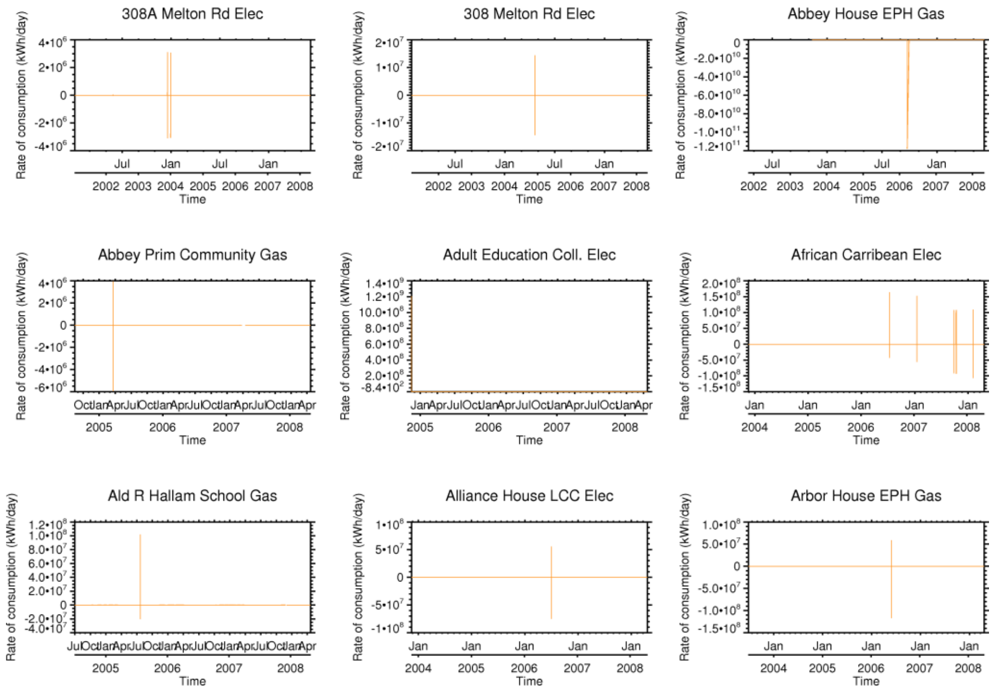


Figure 7.6: Average power calculated for datasets with erroneous readings

Figure 7.6 shows the average power calculated for each pair of meter readings for the same example datasets shown in Figure 7.2. The data errors are so extreme that they entirely dominate the dataset. Variation in the actual consumption data is several orders of magnitude smaller than the size of the extreme values. With such extreme values of r, data errors are often easy to identify by eye.

Spikes generate a pair of extreme values corresponding to the leading and trailing edges of the spike. The magnitude of these extreme values will be determined by the gradient formed by the erroneous reading. They will have opposite signs and their sum will be equal to the actual consumption which occurred during the period between the readings either side of the erroneous reading.

Spikes caused by a corrupted pulse count will generate a roughly equal and opposite pair. If the timestamp or meter reference has been corrupted then the magnitude of the two spikes will be determined by the value of the timestamp relative to the previous and subsequent readings. That is, the erroneous reading will rarely sit half-way between two readings.

Steps in the meter reading data generate a single extreme value for the rate of consumption corresponding to the disjoint in meter readings. As mentioned in section A.2, removing such extreme values from this dataset allows a 'clean' set of meter readings to be calculated.

A.4.2 Maximum normalised residual

The maximum normalised residual provides a measure of the largest absolute deviation from the sample mean expressed in units of standard deviation.

$$\varepsilon_{max} = \frac{\max\limits_{i=1,\dots,N} |r_i - \mu_r|}{\sigma_r}$$

Where r_i is defined as above and the mean, μ_r and standard deviation, σ_r are calculated according to standard equations.

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N} r_i$$

$$\sigma_r = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (r_i - \mu_r)^2}$$

This statistic is referred to in this work as the 'maximum deviation', ε_{max}. It provides a normalised indication of the magnitude of the most extreme value in a dataset. It has no units but can be thought of as being measured in number of standard deviations from the mean. For example, for a dataset where ε_{max}=4.5 the distance between the mean rate of consumption and the largest observed rate of consumption is 4.5 times the standard deviation.

The distribution shown in Figure 7.7 represents a modified version of the maximum deviation calculated across all datasets (in wide bins of 25σ). The sign of the deviation has been preserved in the figure in order to highlight the symmetrical nature of erroneous data. The actual maximum deviation values are always positive and can be seen in Figure 7.9 below.

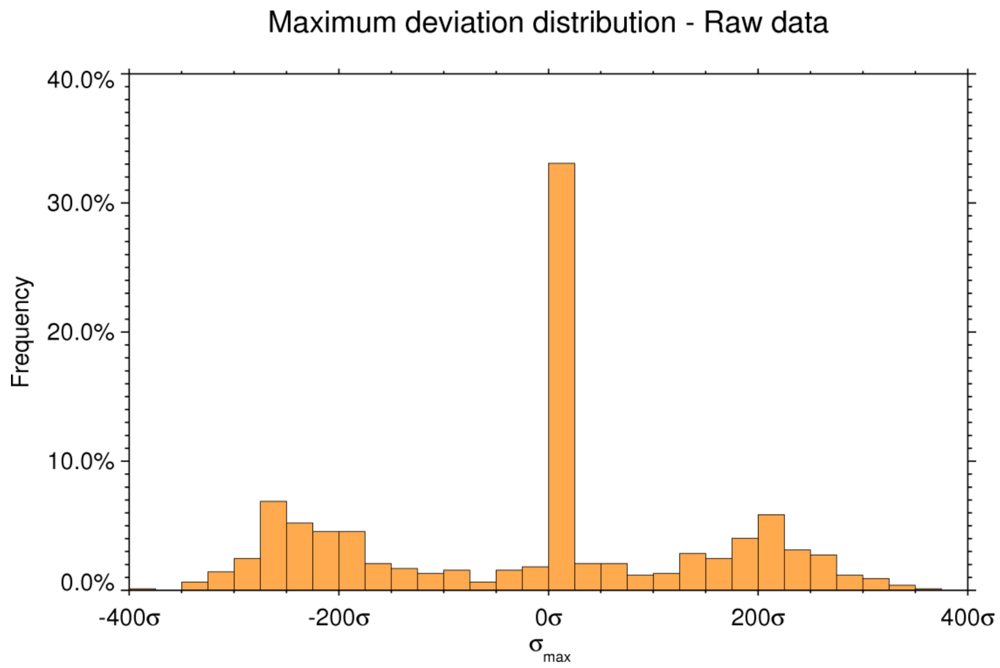


Figure 7.7: Maximum deviation distribution of raw data (with sign preserved)

Figure 7.7 shows two clear features. Over 30% of the datasets have values of ε_{max} between zero and 25σ, also there is a pair of wide, low peaks at around 200σ and -250σ. This is assumed to generally reflect the difference between datasets with and without erroneous readings.

This roughly translates to errors appearing in two-thirds of the datasets, a worrying statistic. However, a dataset only needs a single erroneous point in several tens of thousands to be identified as erroneous. This distribution could be produced with only about 0.001% erroneous data.

It is worth noting that the two distributions either side of zero appear to be roughly normally distributed. This implies that the erroneous data are in some way systematic. This is thought to be related to two factors. Firstly, the maximum deviation statistic is self-dampened. A very large erroneous value will determine the standard deviation and so increase the denominator. Thus very large values are not possible.

For very large outliers the maximum deviation is related to the number of data points in the dataset. It should also be noted that values closer to zero will be under-represented as only the largest error in any one dataset counts. For datasets suffering from two or more erroneous readings only the most deviant reading will be identified.

A.4.3 Grubbs limit

Devising a method for distinguishing between erroneous readings and real consumption data involves identifying an absolute cut off threshold for each dataset above which readings are assumed to be erroneous and below which readings are treated as genuine. This cut off threshold can be determined manually 'by eye' or using a statistical method. A statistical method is preferable for reliability, consistency, replicability and flexibility.

Statistical methods can be used to determine limits for the maximum deviation within which normally distributed data are expected to fall. Any data falling outside of these limits can therefore be considered as coming from a different distribution. More specifically, Grubbs test for outliers (Grubbs 1969) can be used to test the following hypothesis.

H0: There are no outliers in the dataset

Ha: There is at least one outlier in the dataset

Under the assumption of normality, the null hypothesis is rejected at significance level α if the test statistic falls outside the critical region.

$$\sigma_{max} > \frac{N-1}{\sqrt{N}}\sqrt{\frac{t^2_{\frac{a}{N},N-2}}{N-2+t^2_{\frac{a}{N},N-2}}}$$

Where $t_{\alpha/N,\ N-2}$ is the upper critical value of the t-distribution at significance α/N and N-2 degrees of freedom.

The Grubbs test provides a concrete value to use as a threshold for filtering erroneous values. The process for cleaning a dataset of errors is as follows. The Grubbs limit is calculated and each point in the dataset which exceeds the limit is removed. Removing extreme values has a significant impact on the standard deviation so a new limit can be calculated and will necessarily be narrower. The process is applied iteratively until all data fall within the limits.

The datasets used in this research cover a range of N between 1,153 for the shortest dataset and 150,133 for the longest. The Grubbs limit covering this range is shown in Figure 7.8 for four cases (α=0.1, α=0.01, α=0.001, α=0.0001). It is clear that for all datasets in this study, where values of N are well below 200,000 and with α set as low as 0.01% the Grubbs limit remains well below 7σ.

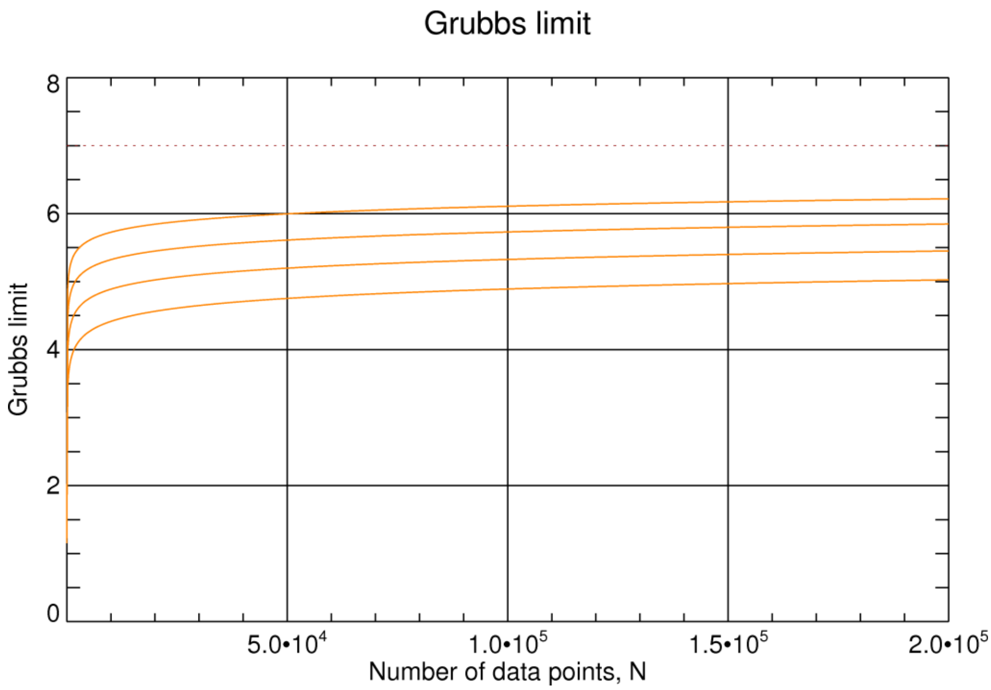


Figure 7.8: Grubbs limit for a range of values for N

Thus, any reading beyond seven standard deviations of the mean (7σ) would be considered a candidate for removal according to the Grubbs test. Under the assumption of normality, applying such a wide limit should remove most erroneous data without 'damaging' the real data values.

However, this 7σ limit applies only when the data are normally distributed. It has been observed that the energy datasets rarely exhibit a normal distribution. Since energy consumption data are not necessarily normally distributed, in many cases this limit may be inappropriate. A pragmatic approach has been taken which attempts to balance the impact of setting the limit too wide and setting the limit too narrow.

A.4.4 Setting a threshold

It is tempting to simply apply the Grubbs limit directly as a means to filter erroneous points. However, it should be made clear that the test assumes the data are normally distributed. This is not always the case with energy consumption data.

Here we will look at the observed values for the maximum deviation as calculated for each of the datasets under analysis before and after the cleaning algorithm is applied. Figure 7.9 shows more detail (0 to 50σ in 1σ bins) of the central peak from Figure 7.7. The absolute figures are shown so all the negative values have switched sign.

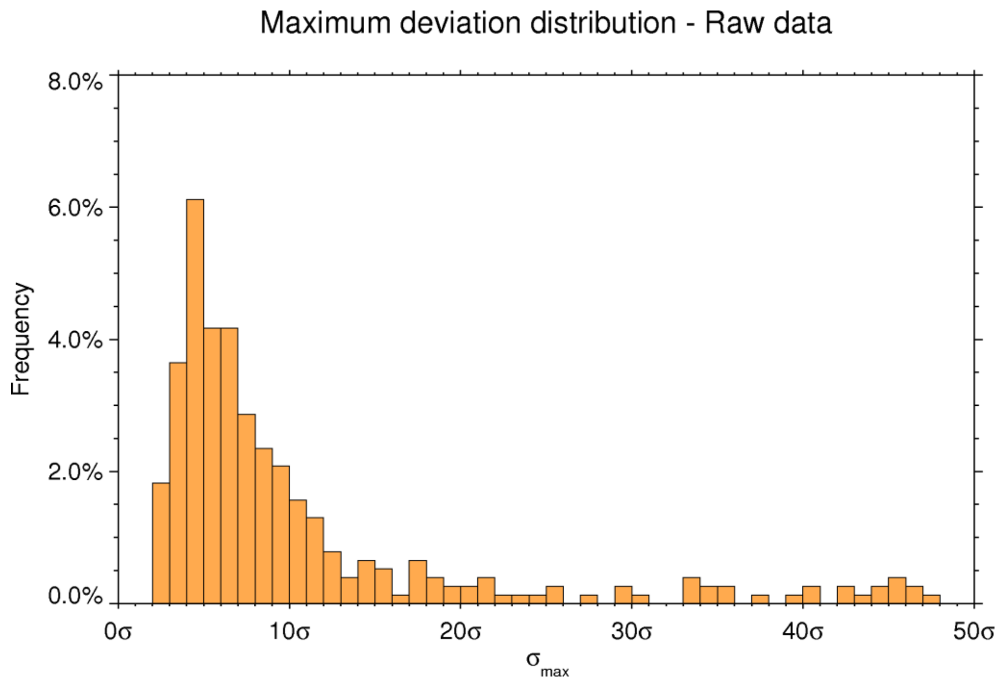


Figure 7.9: The central peak in the maximum deviation distribution

The Grubbs limit has been shown to be below 7σ for all the datasets in this study. Therefore if a dataset is normally distributed it is expected to have a maximum deviation of less than 7σ. If the peak in Figure 7.9 does represent datasets with no errors then we have reason to believe that some of the datasets must have non-normal distributions and higher 'natural' values for maximum deviation than the assumption of normality predicts.

Thus, the Grubbs limit is not applicable. Cleaning the data using this limit would remove valid data for many datasets. The challenge is to determine at what point extremities due to non-normal data give way to erroneous readings.

The data were 'cleaned' using the algorithm described in section A.4.3. The process is represented in Figure 7.10. The algorithm requires a threshold level to determine how aggressively the data are cleaned. A low threshold value will be in danger of cutting into the valid consumption data. A high threshold will be in danger of allowing erroneous points through untouched. The lower bound is set to zero so all negative values are also removed no matter what the threshold.

Since several errors may occur in one dataset, the procedure is applied iteratively until no more outliers are found. Once the most extreme remaining point falls within acceptable boundaries, the dataset can be considered free from both steps and spikes.

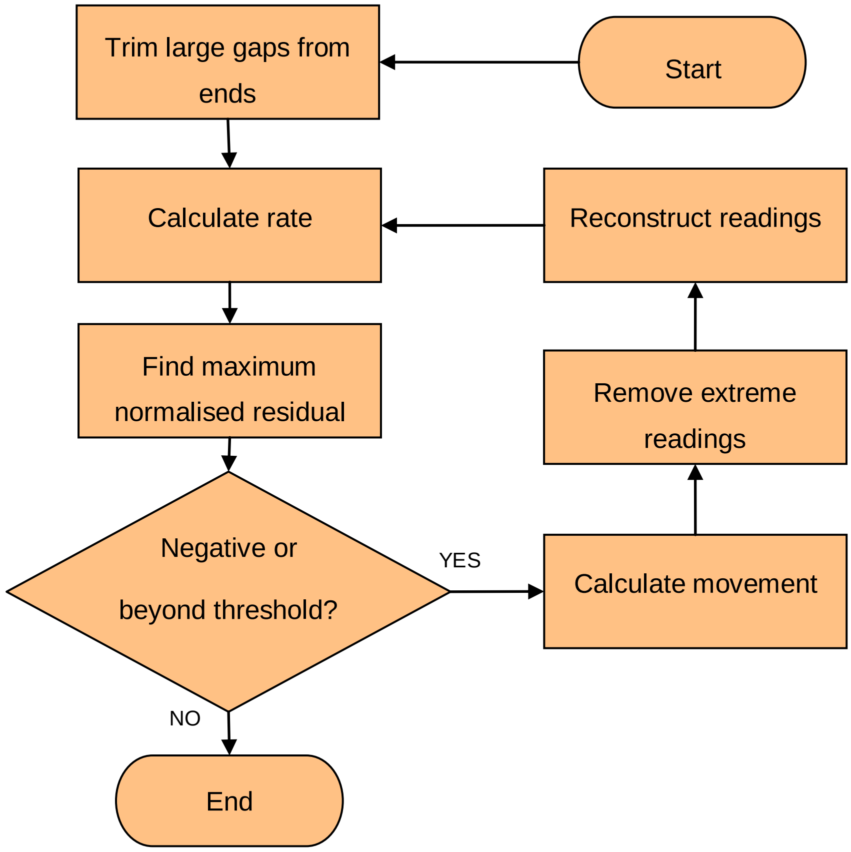


Figure 7.10: Flow chart of data cleaning process

To investigate the impact of varying the threshold, the process was applied to each available dataset using several different threshold values. The highest threshold value was set to 400σ which is above the maximum observed value so only readings which result in negative values will be removed. The lowest threshold was set to 10σ which is well within the peak in Figure 7.9 and so is expected to cause damage to some datasets.

For each dataset, the cleaning process removes zero or more readings to produce a new dataset with extreme values trimmed. The maximum deviation can then be calculated for each of the cleaned datasets. The results of this process are shown in Figure 7.11 and Figure 7.12.

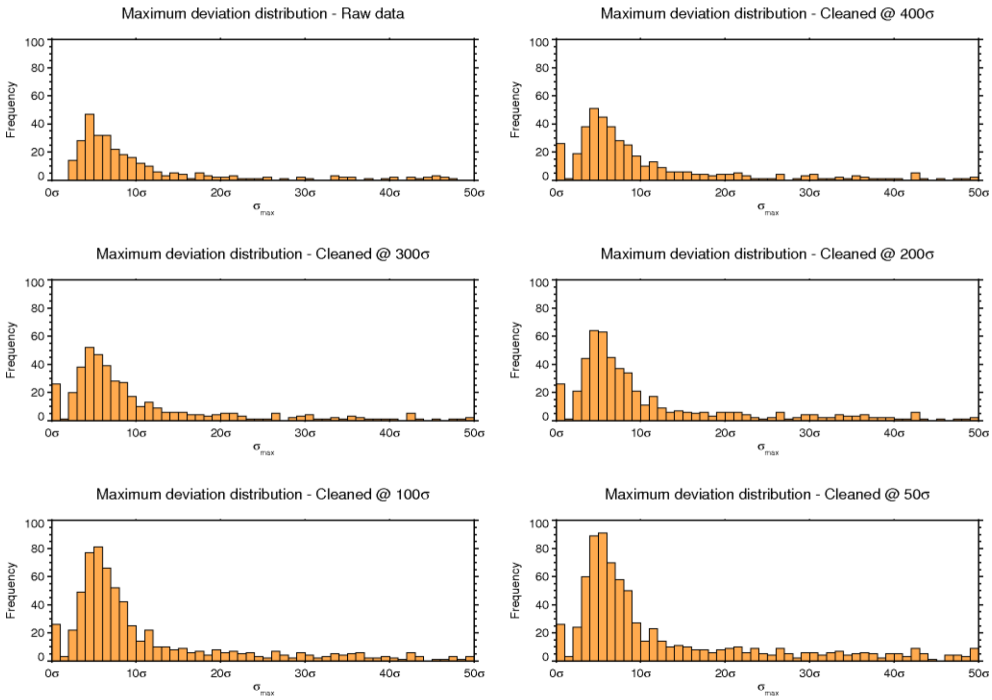


Figure 7.11: Maximum deviation distribution - cleaned @ 400σ – 50σ

The figures show the same range (0σ – 50σ) of the distribution as shown in Figure 7.9. Figure 7.11 shows six distributions, one for the raw data (exactly as in Figure 7.9) and one each for the data cleaned with a threshold of 400σ, 300σ, 200σ, 100σ and 50σ.

The effect of trimming closer and closer to the data in this way fits with the assumption that datasets with large maximum deviations (greater than 50σ) have erroneous readings. The datasets which have readings removed necessarily move to the left of the distribution and for the most part add to the peak indicating that the cleaned data are free from errors and have similar consumption distributions to unaffected datasets.

For those cleaned datasets which add to the tail (above about 15σ) it is assumed at this stage that they originally contained multiple erroneous values, only the largest of which have been removed.

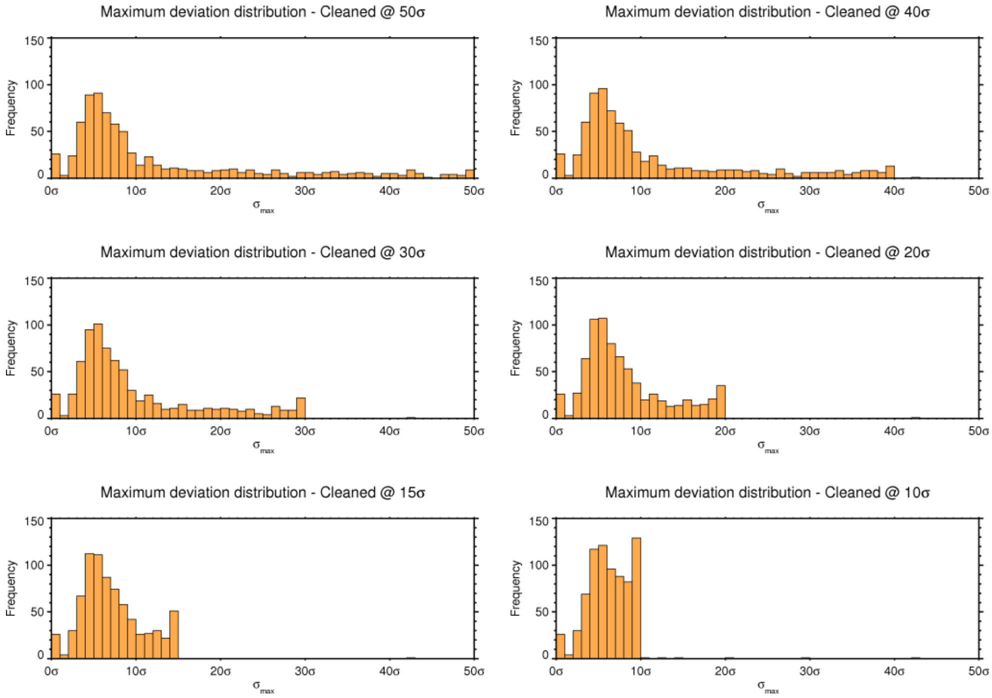


Figure 7.12: Maximum deviation distribution - cleaned @ 50σ – 10σ

Figure 7.12 shows six distributions, one each for the data cleaned with thresholds of 50σ, 40σ, 30σ, 20σ, 15σ and 10σ. The effect of trimming closer and closer to the data in this way broadly follows the same pattern as observed for the wider limits.

However, narrowing the limit beyond 50σ introduces a new feature. There is an increase in the number of datasets for which, after the cleaning process the maximum deviation is very close (in this case within one bin, 1σ) to the threshold value. This can be seen most clearly in the distributions cleaned with thresholds of 20σ and less but is apparent with the wider limits.

The distribution of maximum deviation values across all the raw datasets includes a very long tail. This tail represents datasets with erroneous readings of varying magnitude relative to the standard deviation. Cleaning all the datasets in the way described above will always be expected to result in a small number of datasets where an erroneous reading happens by chance to fall only slightly below the threshold.

As the threshold is reduced and the cleaning process becomes more aggressive there is more chance that extreme but real readings will exceed the threshold. If this is the case then the distribution of consumption would be truncated with the most extreme (but real) values identified as erroneous. Under these circumstances the new maximum deviation would be expected to be very close to the threshold value. It is this additional effect that is being seen in Figure 7.12.

A threshold of 30σ has been used in this work. This was chosen as a compromise between allowing large errors to pass through the filter and damaging data by setting the threshold too low.

In addition to the cleaning process datasets with a resultant maximum deviation greater than 20σ were flagged as being in the ‘danger zone’ where they may still potentially contain errors and also are the most likely to have had their consumption distribution truncated. Flagged datasets will be identified as such when the data are interpreted.

A.5 Discussion

The data cleaning algorithm described in this appendix raises certain questions as does the assessment of the quality of the source data for this work.

A.5.1 Cause of errors

It is very difficult to piece together the precise cause of missing and erroneous data from the information available. What can be said with confidence is that data errors have been introduced prior to the data being imported into the central database since they are present in the raw data. This means they cannot be eradicated by going back to the source.

Missing data and spikes are likely introduced during communication, it is understood that the systems use a very simple checksum to validate readings. It is unfortunate that these problems exist since it is likely they could have been avoided by the use of more rigorous data validation such as a position dependent checksum to confirm data integrity.

Discussions with the system designers indicate that steps may be related to the replacement of faulty metering hardware without properly resetting the pulse count. In this scenario the replacement hardware has an internal counter that is different from the expected value and this difference feeds through the system to create a step. Again, this is unfortunate since it is a problem that could have been avoided by the introduction of a proper procedure for swapping hardware in this way.

The precise cause of erroneous data is unknown, spikes can be thought of as randomised bits in the meter reading record. Steps are assumed to be the result of randomisation at a different stage in the data processing which causes the outlier to be absorbed into the consumption dataset rather than the pulse count.

This would have some implications, for example if the pulse count is stored as an unsigned 32 bit integer then the range of potential values is between zero (20-1) and 4,294,967,295 (232-1). Any randomisation of the individual bits of data which make up the digital (binary) representation of that number cannot lead to a value beyond these fundamental limits. This limit is likely to be well beyond the expected values of the actual meter readings. Thus, if readings are truly randomised, they should be distinguishable from the vast majority of errors by their scale alone.

Random errors in the timestamp fields would be far less common since, depending on the underlying storage format, the range of allowable values is likely to be less. Any invalid dates or times (e.g. months greater than 12 or minutes greater than 60) are rejected by the data import software and become simple missing data.

Errors in the meter reference field would need to be very specific to be valid. thus it is expected that they are very rare if they occur at all. However, it is possible that meters with similar references could be confused and mixed by random errors. Again, these conclusions depend greatly on the underlying data format during the journey from meter to database and on the specific cause of data corruption.

A.5.2 Effectiveness

The cleaning process described above is complex and though based on a statistical treatment is not entirely robust. However, the approach is consistent and replicable and can be shown broadly to offer a reasonable compromise between an overly aggressive approach and an overly permissive approach.

In the end, this processing was unavoidable since the dataset was effectively unusable in its original form. This processing has led to the dataset being entirely satisfactory for the analyses conducted in this work.